On Recovery Guarantees for One-Bit Compressed Sensing on Manifolds

Mark A. Iwen[1] · Felix Krahmer[2] · Sara Krause-Solberg[3] · Johannes Maly[4]

Abstract

This paper studies the problem of recovering a signal from one-bit compressed sensing measurements under a manifold model; that is, assuming that the signal lies on or near a manifold of low intrinsic dimension. We provide a convex recovery method based on the Geometric Multi-Resolution Analysis and prove recovery guarantees with a near-optimal scaling in the intrinsic manifold dimension. Our method is the first tractable algorithm with such guarantees for this setting. The results are complemented by numerical experiments confirming the validity of our approach.

Editor in Charge: Kenneth Clarkson

Mark A. Iwen
iwenmark@msu.edu

Felix Krahmer
felix.krahmer@tum.de

Sara Krause-Solberg
sara.krause-solberg@desy.de

Johannes Maly
maly@mathc.rwth-aachen.de

[1] Department of Mathematics, and Department of Computational Mathematics, Science, and Engineering (CMSE), Michigan State University, East Lansing, MI 48824, USA

[2] Department of Mathematics, Technical University of Munich, 85748 Garching, Germany

[3] HIP Helmholtz Imaging Platform, DESY, 22607 Hamburg, Germany

[4] Department of Mathematics, RWTH University Aachen, 52062 Aachen, Germany

⚇ Springer

1 Introduction

Linear inverse problems are ubiquitous in many applications in science and engineering. Starting with the seminal works of Candès et al. [10], as well as Donoho [15], a new paradigm in their analysis became an active area of research in the last decades. Namely, rather than considering the linear model as entirely given by the application, one seeks to actively choose remaining degrees of freedom, often using a randomized strategy, to make the problem less ill-posed. This approach gave rise to a number of recovery guarantees for random linear measurement models under structural data assumptions. The first works considered the recovery of sparse signals; subsequent works analyzed more general union-of-subspaces models [18] and the recovery of low-rank matrices [44], a model that can also be employed when studying phaseless reconstruction problems [11] or bilinear inverse problems [1].

Another line of works following this approach studies manifold models. That is, one assumes that the structural constraints are given by (unions of finitely many) manifolds. While this model is considerably richer than say sparsity, its rather general formulation makes a unified study, at least in some cases, somewhat more involved. The first work to study random linear projections of smooth manifold was [5], the authors show that Gaussian linear dimension reductions typically preserve the geometric structure. In [29], these results are refined and complemented by a recovery algorithm, which is based on the concept of the Geometric Multi-Resolution Analysis as introduced in [3] (cf. Sect. 2.1 below). These results were again substantially improved in [17]; these latest results no longer explicitly depend on the ambient dimension.

Arguably, working with manifold models is better adapted to real-world data than sparsity and hence may allow one to work with smaller embedding dimensions. For that, however, other practical issues need to be considered as well. In particular, to our knowledge there are almost no works to date that study the effects of quantization, i.e., representing the measurements using only a finite number of bits (the only remotely connected work that we are aware of is [39], but this paper does not consider dimension reduction and exclusively focuses on the special case of Grassmann manifolds).

For sparse signal models, in contrast, quantization of subsampled random measurements is an active area of research. On the one hand, a number of works considered the scenario of memoryless scalar quantization, that is, each of the measurements is quantized independently. In particular, the special case of representing each measurement only by a single bit, its sign—often referred to as *one-bit compressed sensing*—has received considerable attention. In [31], it was shown that one-bit compressed sensing with Gaussian measurements approximately preserves the geometry, and a heuristic recovery scheme was presented. In [42], recovery guarantees for a linear method, again with Gaussian measurements, were derived. Subsequently, these results were generalized to sub-Gaussian measurements [2], and partial random circulant measurements [14]. In [41], the authors provided a recovery procedure for noisy one-bit Gaussian measurements which provably works on more general signal sets (essentially arbitrary subsets of the Euclidean ball). This procedure, however, becomes NP-hard as soon as the signal set is non-convex, a common property of manifolds.

Another line of works studied the so-called feedback quantizers, that is, the bit sequence encoding the measurements is computed using a recursive procedure. These

works adapt the Sigma-Delta modulation approach originally introduced in the context of bandlimited signals [22,40] and later generalized to frame expansions [6,7] to the sparse recovery framework. A first such approach was introduced and analyzed for Gaussian measurements in [23]; subsequent works generalize the results to sub-Gaussian random measurements [20,33]. Recovery guarantees for a more stable reconstruction scheme based on convex optimization were proven for sub-Gaussian measurements in [45] and extended to partial random circulant matrices in [21]. For more details on the mathematical analysis available for different scenarios, we refer the reader to the overview chapter [9].

In this paper, we focus on the MSQ approach and leave the study of Sigma-Delta quantizers under manifold model assumptions for future work.

1.1 Contribution

We provide the first tractable one-bit compressed sensing algorithm for signals which are well approximated by manifold models. It is simple to implement and comes with error bounds that basically match the state-of-the-art recovery guarantees in [41]. In contrast to the minimization problem introduced in [41], which does not come with a minimization algorithm, our approach always admits a convex formulation and hence allows for tractable recovery. Our approach is based on the Geometric Multi-Resolution Analysis (GMRA) introduced in [3], and hence combines the approaches of [29] with the general results for one-bit quantized linear measurements provided in [41,43].

1.2 Outline

We begin by a detailed description of our problem in Sect. 2 and fix notation for the rest of the paper. The section also includes a complete axiomatic definition of GMRA. Section 3 states our main results. The proofs can be found in Sect. 4. In Sect. 5, we present some numerical experiments testing the recovery in practice and conclude with Sect. 6. Technical parts of the proofs as well as adaption of the results to GMRAs from random samples are deferred to the appendix.

2 Problem Formulation, Notation, and Setup

The problem we address is the following. We consider a given union of low-dimensional manifolds (i.e., signal class) $\mathscr{M}$ of intrinsic dimension d that is a subset of the unit sphere $\mathbb{S}^{D-1}$ of a higher dimensional space $\mathbb{R}^D$, $d \ll D$. Furthermore, let us imagine that we do not know $\mathscr{M}$ perfectly, and so instead we only have approximate information about $\mathscr{M}$ represented in terms of a structured dictionary model $\mathcal{D}$ for the manifold. Our goal is now to recover an unknown signal $\mathbf{x} \in \mathscr{M}$ from m one-bit measurements

$$\mathbf{y} = \text{sign } A\mathbf{x}, \tag{1}$$

where $A \in \mathbb{R}^{m \times D}$ has Gaussian i.i.d. entries of variance $1/\sqrt{m}$, using as few measurements, m, as possible. Each single measurement sign $\langle \mathbf{a}_i, \mathbf{x} \rangle$ can be interpreted as the random hyperplane $\{\mathbf{z} \in \mathbb{R}^D : \langle \mathbf{a}_i, \mathbf{z} \rangle = 0\}$ tessellating the sphere (cf. Fig. 1a). In order to succeed using only $m \ll D$ such one-bit measurements we will use the fact that $\mathbf{x}$ is approximately sparse in our (highly coherent, but structured) dictionary $\mathcal{D}$ for $\mathcal{M}$ provides structural constraints for signal $\mathbf{x}$ to be recovered. Thus the setup connects to recent generalizations of the quantized compressed sensing problem [41] which we will exploit in our proof.

## 2.1 GMRA Approximations to $\mathcal{M} \subset \mathbb{R}^D$ and Two Notions of Complexity

Clearly, the solution to this problem depends on what kind of representation, $\mathcal{D}$, of the manifold $\mathcal{M}$ we have access to. In this paper, we consider the scenario where the dictionary for the manifold is provided by a Geometric Multi-Resolution Analysis (GMRA) approximation to $\mathcal{M}$ [3] (cf. Fig. 1b). We will mainly work with GMRA approximations of $\mathcal{M}$ characterized by the axiomatic Definition 2.1 below, but we also consider the case of a GMRA approximation based on random samples from $\mathcal{M}$ (see Sect. 3 and Appendices D and E for more details).

As one might expect, the complexity and structure of the GMRA-based dictionary for $\mathcal{M}$ will depend on the complexity of $\mathcal{M}$ itself. In this paper, we will work with two different measures of a set's complexity: (i) the set's *Gaussian width* and (ii) the notion of the *reach* of the set [19]. The Gaussian width of a set $\mathcal{M} \subset \mathbb{R}^D$ is defined by

$$w(\mathcal{M}) := \mathbb{E}\left[\sup_{\mathbf{z} \in \mathcal{M}} \langle \mathbf{g}, \mathbf{z} \rangle\right]$$
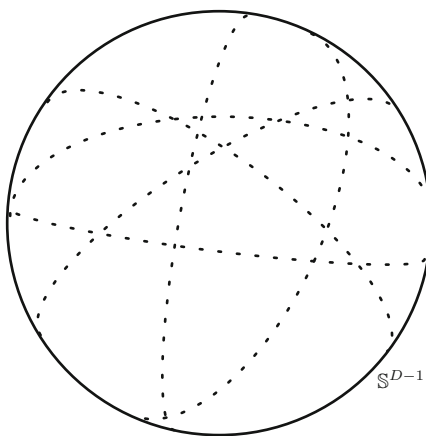
(a) Tessellation of the sphere by random hyperplanes.
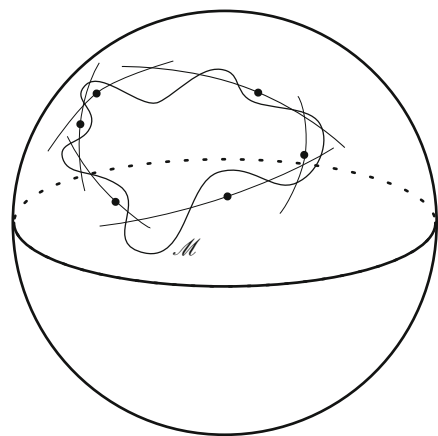
(b) Submanifold $\mathcal{M}$ of $\mathbb{S}^{D-1}$ and one level of GMRA.

Fig. 1 One-bit measurements and GMRA

where $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, I_D)$. Properties of this quantity are discussed in Sect. 4.1. The notion of reach is, in contrast, more obviously linked to the geometry of $\mathcal{M}$ and requires a couple of additional definitions before it can be defined formally.

The first of these definitions is the *tube of radius r* around a given subset $\mathcal{M} \subset \mathbb{R}^D$, which is the D-dimensional superset of $\mathcal{M}$ consisting of all the points in $\mathbb{R}^D$ that are within Euclidean distance $r \geq 0$ from $\mathcal{M} \subset \mathbb{R}^D$,

$$\text{tube}_r \mathcal{M} := \left\{ \mathbf{x} \in \mathbb{R}^D : \inf_{\mathbf{y} \in \mathcal{M}} \|\mathbf{x} - \mathbf{y}\|_2 \leq r \right\}.$$

The domain of the nearest neighbor projection onto the closure of $\mathcal{M}$ is also needed, and is denoted by

$$D(\mathcal{M}) := \left\{ \mathbf{x} \in \mathbb{R}^D : \exists! \, \mathbf{y} \in \overline{\mathcal{M}} \text{ such that } \|\mathbf{x} - \mathbf{y}\|_2 = \inf_{\mathbf{z} \in \mathcal{M}} \|\mathbf{x} - \mathbf{z}\|_2 \right\}.$$

Finally, the reach of the set $\mathcal{M} \subset \mathbb{R}^D$ is simply defined to be the smallest distance r around $\mathcal{M}$ for which the nearest neighbor projection onto the closure of $\mathcal{M}$ is no longer well defined. Equivalently,

$$\text{reach}\, \mathcal{M} := \sup \{r \geq 0 : \text{tube}_r \mathcal{M} \subseteq D(\mathcal{M})\}.$$

Given this definition one can see, e.g., that the reach of any $d < D$ dimensional sphere of radius r in $\mathbb{R}^D$ is always r, and that the reach of any $d \leq D$ dimensional convex subset of $\mathbb{R}^D$ is always ∞.

Definition 2.1 (GMRA approximation to $\mathcal{M}$ [29]) Let $J \in \mathbb{N}$ and $K_0, K_1, \ldots, K_J \in \mathbb{N}$. Then a *Geometric Multi-Resolution Analysis (GMRA) approximation* of $\mathcal{M}$ is a collection $\{(\mathcal{C}_j, \mathcal{P}_j)\}$, $j \in [J] := \{0, \ldots, J\}$, of sets $\mathcal{C}_j = \{\mathbf{c}_{j,k}\}_{k=1}^{K_j} \subset \mathbb{R}^D$ of centers and

$$\mathcal{P}_j = \left\{ \mathbb{P}_{j,k} : \mathbb{R}^D \to \mathbb{R}^D \mid k \in [K_j] \right\}$$

of affine projectors that approximate $\mathcal{M}$ at scale j, such that the following assumptions (i)–(iii) hold.

(i) Affine projections: Every $\mathbb{P}_{j,k} \in \mathcal{P}_j$ has both an associated center $\mathbf{c}_{j,k} \in \mathcal{C}_j$ and an orthogonal matrix $\Phi_{j,k} \in \mathbb{R}^{d \times D}$ such that

$$\mathbb{P}_{j,k}(\mathbf{z}) = \Phi_{j,k}^T \Phi_{j,k}(\mathbf{z} - \mathbf{c}_{j,k}) + \mathbf{c}_{j,k},$$

i.e., $\mathbb{P}_{j,k}$ is the projector onto some affine d-dimensional linear subspace $P_{j,k}$ containing $\mathbf{c}_{j,k}$.

(ii) Dyadic structure: The number of centers at each level is bounded by $|\mathcal{C}_j| = K_j \leq C_{\mathcal{C}} 2^{dj}$ for an absolute constant $C_{\mathcal{C}} \geq 1$. There exist $C_1 > 0$ and $C_2 \in (0, 1]$, such that following conditions are satisfied:

(a) $K_j \leq K_{j+1}$ for all $j \in [J - 1]$.

(b) $\|\mathbf{c}_{j,k_1} - \mathbf{c}_{j,k_2}\|_2 > C_1 \cdot 2^{-j}$ for all $j \in [J]$ and $k_1 \neq k_2 \in [K_j]$.

(c) For each $j \in [J] \setminus \{0\}$ there exists a parent function $p_j \colon [K_j] \to [K_{j-1}]$ with

$$\|\mathbf{c}_{j,k} - \mathbf{c}_{j-1,p_j(k)}\|_2 \leq C_2 \cdot \min_{k' \in [K_{j-1}] \setminus \{p_j(k)\}} \|\mathbf{c}_{j,k} - \mathbf{c}_{j-1,k'}\|_2.$$

(iii) Multiscale approximation: The projectors in $\mathscr{P}_j$ approximate $\mathscr{M}$ at scale j, i.e., when $\mathscr{M}$ is sufficiently smooth the affine spaces $P_{j,k}$ locally approximate $\mathscr{M}$ pointwise with error $\mathscr{O}(2^{-2j})$. More precisely:

(a) There exists $j_0 \in [J-1]$ such that $\mathbf{c}_{j,k} \in \text{tube}_{C_1 2^{-j-2}}(\mathscr{M})$ for all $j > j_0 \geq 1$ and $k \in [K_j]$.

(b) For each $j \in [J]$ and $\mathbf{z} \in \mathbb{R}^D$ let $\mathbf{c}_{j,k_j(\mathbf{z})}$ be one of the centers closest to $\mathbf{z}$, i.e.,

$$k_j(\mathbf{z}) \in \operatorname*{arg\,min}_{k \in [K_j]} \|\mathbf{z} - \mathbf{c}_{j,k}\|_2. \tag{2}$$

Then, for each $\mathbf{z} \in \mathscr{M}$ there exists a constant $C_{\mathbf{z}} > 0$ such that

$$\|\mathbf{z} - \mathbb{P}_{j,k_j(\mathbf{z})}(\mathbf{z})\|_2 \leq C_{\mathbf{z}} \cdot 2^{-2j}$$

for all $j \in [J]$. Moreover, for each $\mathbf{z} \in \mathscr{M}$ there exists $\tilde{C}_{\mathbf{z}} > 0$ such that

$$\|\mathbf{z} - \mathbb{P}_{j,k'}(\mathbf{z})\|_2 \leq \tilde{C}_{\mathbf{z}} \cdot 2^{-j}$$

for all $j \in [J]$ and $k' \in [K_j]$ satisfying

$$\|\mathbf{z} - \mathbf{c}_{j,k'}\|_2 \leq 16 \cdot \max \left\{ \|\mathbf{z} - \mathbf{c}_{j,k_j(\mathbf{z})}\|_2, C_1 \cdot 2^{-j-1} \right\}.$$

Remark 2.2 By property (i), GMRA approximation represents $\mathscr{M}$ as a combination of several anchor points (the centers $\mathbf{c}_{j,k}$) and corresponding low-dimensional affine spaces $P_{j,k}$. The levels j control the accuracy of the approximation. The centers are organized in a tree-like structure as stated in property (ii). Property (iii) then characterizes approximation criteria to be fulfilled on different refinement levels. Note that centers do not have to lie on $\mathscr{M}$ (compare Fig. 1b) but their distance to $\mathscr{M}$ is controlled by property (iii.a). If the centers form a maximal 2^{-j} packing of a smooth manifold $\mathscr{M}$ at each scale j or if the GMRA is constructed from manifold samples as discussed in [38] (cf. Appendix E), the constants C_1 and $\tilde{C}_{\mathbf{z}}$ are in fact bounded by absolute constants which will become important later on, cf. Remark 3.2.

2.2 Additional Notation

Let us now fix some additional notation. Throughout the remainder of this paper we will work with several different metrics. Perhaps most importantly, we will quantify
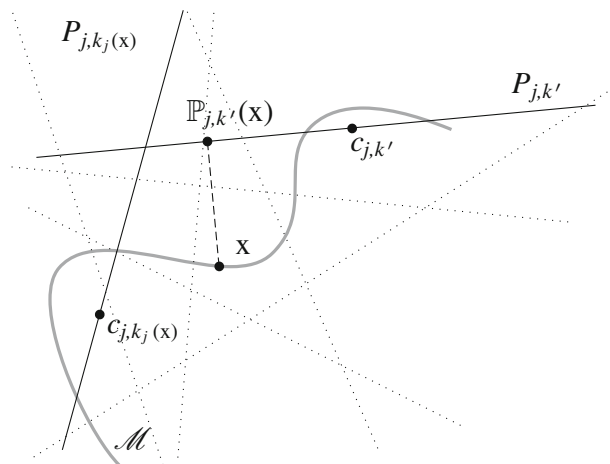
Fig. 2 The closest center $\mathbf{c}_{j,k_j(\mathbf{x})}$ is not identified by measurements. Dotted lines represent one-bit hyperplanes

the distance between two points $\mathbf{z}, \mathbf{z}' \in \mathbb{R}^D$ with respect to their one-bit measurements by

$$d_A(\mathbf{z}, \mathbf{z}') := \frac{d_H(\text{sign } A\mathbf{z}, \text{sign } A\mathbf{z}')}{m},$$

where d_H counts the number of differing entries between the two sign patterns (i.e., $d_A(\mathbf{z}, \mathbf{z}')$ is the normalized Hamming distance between the signs of $A\mathbf{z}$ and $A\mathbf{z}'$). Furthermore, let $\mathbb{P}_{\mathbb{S}}$ denote orthogonal projection onto the unit sphere $\mathbb{S}^{D-1}$, and more generally let $\mathbb{P}_K$ denote orthogonal (i.e., nearest neighbor) projection onto the closure of an arbitrary set $K \subset \mathbb{R}^D$ wherever it is defined. Then, for all $\mathbf{z}, \mathbf{z}' \in \mathbb{R}^D$ we will denote by $d_G(\mathbf{z}, \mathbf{z}') = d_G(\mathbb{P}_{\mathbb{S}}(\mathbf{z}), \mathbb{P}_{\mathbb{S}}(\mathbf{z}'))$ the geodesic distance between $\mathbb{P}_{\mathbb{S}}(\mathbf{z})$ and $P_{\mathbb{S}}(\mathbf{z}')$ on $\mathbb{S}^{D-1}$ normalized to fulfill $d_G(\mathbf{z}'', -\mathbf{z}'') = 1$ for all $\mathbf{z}'' \in \mathbb{R}^D$.

Herein the Euclidian ball with center $\mathbf{z}$ and radius r is denoted by $\mathscr{B}(\mathbf{z}, r)$. In addition, the *scale-j GMRA approximation to $\mathscr{M}$*,

$$\mathscr{M}_j := \{\mathbb{P}_{j,k_j(\mathbf{z})}(\mathbf{z}) : \mathbf{z} \in \mathscr{B}(\mathbf{0}, 2)\} \cap \mathscr{B}(\mathbf{0}, 2),$$

will refer to the portions of the affine subspaces introduced in Definition 2.1 for each fixed j, which are potentially relevant as approximations to some portion of $\mathscr{M} \subset \mathbb{S}^{D-1}$. To prevent the $\mathscr{M}_j$ above from being empty we will further assume in our results that we only use scales $j > j_0$ large enough to guarantee that $\text{tube}_{C_1 2^{-j-2}}(\mathscr{M}) \subset \mathscr{B}(\mathbf{0}, 2)$. Hence we will have $\mathbf{c}_{j,k} \in \mathscr{B}(\mathbf{0}, 2)$ for all $k \in K_j$, and so $\mathscr{C}_j \subset \mathscr{M}_j$. This further guarantees that no sets $P_{j,k} \cap \mathscr{B}(\mathbf{0}, 2)$ are empty, and that $P_{j,k} \cap \mathscr{B}(\mathbf{0}, 2) \subset \mathscr{M}_j$ for all $k \in K_j$.

Finally, we write $a \gtrsim b$ if $a \geq Cb$ for some constant $C > 0$. The diameter of a set $K \subset \mathbb{R}^D$ will be denoted by $\text{diam } K := \sup_{\mathbf{z}, \mathbf{z}' \in K} \|\mathbf{z} - \mathbf{z}'\|_2$, where $\|\cdot\|_2$ is the Euclidean norm. We use $\text{dist}(A, B) = \inf_{\mathbf{a} \in A, \mathbf{b} \in B} \|\mathbf{a} - \mathbf{b}\|_2$ for the distance of

two sets $A, B \subset \mathbb{R}^D$ and, by an abuse of notation, $\text{dist}(\mathbf{0}, A) = \inf_{\mathbf{a} \in A} \|\mathbf{a}\|_2$. The operator norm of a matrix $A \in \mathbb{R}^{n_1 \times n_2}$ is denoted by $\|A\| = \sup_{\mathbf{x} \in \mathbb{R}^{n_2}, \|\mathbf{x}\|_2 \leq 1} \|A\mathbf{x}\|_2$. We will write $\mathcal{N}(K, \varepsilon)$ to denote the Euclidian covering number of a set $K \subset \mathbb{R}^D$ by Euclidean balls of radius ε (i.e., $\mathcal{N}(K, \varepsilon)$ is the minimum number of ε-balls that are required to cover K). And, the operators $\lfloor r \rfloor$ (resp. $\lceil r \rceil$) return the closest integer smaller (resp. larger) than $r \in \mathbb{R}$.

2.3 The Proposed Computational Approach

Combining prior GMRA-based compressed sensing results [29] with the one-bit results of Plan and Vershynin in [41] suggests the following strategy for recovering an unknown $\mathbf{x} \in \mathcal{M}$ from the measurements given in (1): First, choose a center $\mathbf{c}_{j,k'}$ whose one-bit measurements agree with as many one-bit measurements of $\mathbf{x}$ as possible. Due to the varying shape of the tessellation cells this is not an optimal choice in general (see Fig. 2). Nevertheless, one can expect $P_{j,k'}$ to be a good approximation to $\mathcal{M}$ near $\mathbf{x}$. Thus, in the second step a modified version of Plan and Vershynin's noisy one-bit recovery method using $P_{j,k'}$ should yield an approximation of $\mathbb{P}_{j,k'}(\mathbf{x})$ which is close to $\mathbf{x}$.[1] See OMS-simple for pseudocode.

Algorithm OMS-simple: OnebitManifoldSensing – Simple Version

INPUT: A, $\mathbf{y}$, GMRA of $\mathcal{M}$, refinement level j, R

  I. Identify a center $\mathbf{c}_{j,k'}$ close to $\mathbf{x}$ via

$$\mathbf{c}_{j,k'} \in \arg \min_{\mathbf{c}_{j,k} \in \mathscr{C}_j} d_{\mathrm{H}}(\text{sign } A\mathbf{c}_{j,k}, \mathbf{y}), \tag{3}$$

     where d_{H} is the Hamming distance, i.e., $d_{\mathrm{H}}(\mathbf{z}, \mathbf{z}') := |\{l : z_l \neq z_l'\}|$.
     If $d_{\mathrm{H}}(\text{sign } A\mathbf{c}_{j,k'}, \mathbf{y}) = 0$, directly choose $\mathbf{x}^* = \mathbf{c}_{j,k'}$ and omit II.
  II. If there is no center in the same cell as $\mathbf{x}$ (as in Fig. 2), solve a noisy one-bit recovery problem as in [41], i.e.,

$$\mathbf{x}^* = \arg \min_{\mathbf{z} \in \mathbb{R}^D} \sum_{l=1}^{m} (-y_l)\langle \mathbf{a}_l, \mathbf{z} \rangle \quad \text{subject to } \mathbf{z} = \mathbb{P}_{j,k'}(\mathbf{z}) \text{ and } \|\mathbf{z}\|_2 \leq R \tag{4}$$

     where R is a suitable parameter.

Remark 2.3 The minimization in (3) can be efficiently calculated by exploiting tree structures in $\mathscr{C}_j$. Numerical experiments (see Sect. 5) suggest this strategy to yield adequate approximation for the center $\mathbf{c}_{j,k_j(\mathbf{x})}$ in (2), while being considerably faster (we observed differences in runtime up to a factor of 10).

---

[1] Note that in this second step the given measurements $\mathbf{y}$ of $\mathbf{x}$ are interpreted as being noisy measurements of $\mathbb{P}_{j,k'}(\mathbf{x})$.
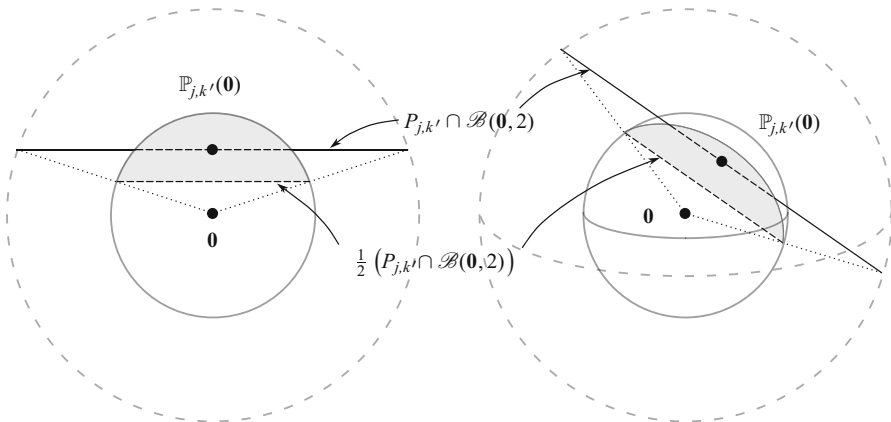
Fig. 3 Two views of an admissible set conv $\mathbb{P}_{\mathbb{S}}(P_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2))$ from (5) if $\|\mathbf{c}\|_2 = \|\mathbb{P}_{j,k'}(\mathbf{0})\|_2 < 1$

Though simple to understand, the constraints in (4) have two issues that we need to address: First, in some cases the minimization problem (4) empirically exhibits suboptimal recovery performance (see Sect. 5.1 for details). Second, the parameter R in (4) is unknown a priori (i.e., OMS-simple requires parameter tuning, making it less practical than one might like). Indeed, our analysis shows that making an optimal choice for R in OMS-simple requires a priori knowledge about $\|\mathbb{P}_{j,k'}(\mathbf{x})\|_2$ which is only approximately known in advance.

To address this issue, we will modify the constraints in (4) and instead minimize over the convex hull of the nearest neighbor projection of $P_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2)$ onto $\mathbb{S}^{D-1}$,

$$\text{conv } \mathbb{P}_{\mathbb{S}}(P_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2)),$$

to remove the R dependence. If $\mathbf{0} \in P_{j,k'}$ one has conv $\mathbb{P}_{\mathbb{S}}(P_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2)) = P_{j,k'} \cap \mathscr{B}(\mathbf{0}, 1)$. If $\mathbf{0} \notin P_{j,k'}$ the set conv $\mathbb{P}_{\mathbb{S}}(P_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2))$ is described by the following set of convex constraints which are straightforward to implement in practice. Denote by $\mathbb{P}_{\mathbf{c}}$ the projection onto the vector $\mathbf{c} = \mathbb{P}_{j,k'}(\mathbf{0})$. Then

$$\mathbf{z} \in \text{conv } \mathbb{P}_{\mathbb{S}}(P_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2)) \quad \Longleftrightarrow \quad \begin{cases} \|\mathbf{z}\|_2 \leq 1, \\ \Phi_{j,k'}^T \Phi_{j,k'} \mathbf{z} + \mathbb{P}_{\mathbf{c}}(\mathbf{z}) = \mathbf{z}, \\ \langle \mathbf{z}, \mathbf{c} \rangle \geq \|\mathbf{c}\|_2^2/2. \end{cases} \tag{5}$$

The first two conditions above restrict $\mathbf{z}$ to $\mathscr{B}(\mathbf{0}, 1)$ and span $P_{j,k'}$, respectively. The third condition then removes all points that are too close to the origin (see Fig. 3). A rigorous proof of equivalence can be found in Appendix A.

Our analysis uses that the noisy one-bit recovery results of Plan and Vershynin apply to arbitrary subsets of the unit ball $\mathscr{B}(\mathbf{0}, 1) \subset \mathbb{R}^D$ which will allow us to adapt our recovery approach. Replacing the constraints in (4) with those in (5) we obtain the following modified recovery approach, OMS.

Algorithm OMS: OnebitManifoldSensing

INPUT: A, $\mathbf{y}$, GMRA of $\mathcal{M}$, refinement level j

I. Identify a center $\mathbf{c}_{j,k'}$ close to $\mathbf{x}$ via

$$\mathbf{c}_{j,k'} \in \underset{\mathbf{c}_{j,k} \in \mathscr{C}_j}{\arg\min}\, d_{\mathrm{H}}(\operatorname{sign} A\mathbf{c}_{j,k}, \mathbf{y}). \tag{6}$$

where d_{H} is the Hamming distance, i.e., $d_{\mathrm{H}}(\mathbf{z}, \mathbf{z}') := |\{l : z_l \neq z_l'\}|$.
If $d_{\mathrm{H}}(\operatorname{sign}(A\mathbf{c}_{j,k'}), \mathbf{y}) = 0$, directly choose $\mathbf{x}^* = \mathbf{c}_{j,k'}$ and omit II.

II. If there is no center lying in the same cell as $\mathbf{x}$ (see Fig. 2), recover the projection of $\mathbf{x}$ onto $P_{j,k'}$, i.e., $\mathbb{P}_{j,k'}(\mathbf{x})$. To do so solve the convex optimization

$$\mathbf{x}^* = \underset{\mathbf{z} \in \mathbb{R}^D}{\arg\min} \sum_{l=1}^{m} (-y_l)\langle \mathbf{a}_l, \mathbf{z} \rangle \quad \text{subject to } \mathbf{z} \in \operatorname{conv} \mathbb{P}_{\mathbb{S}}(P_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2)). \tag{7}$$

As we shall see, theoretical error bounds for both OMS-simple and OMS can be obtained by nearly the same analysis despite their differences.

3 Main Results

In this section, we present the main results of our work, namely that both OMS-simple and OMS approximate a signal on $\mathcal{M}$ to arbitrary precision with a near-optimal number of measurements. More precisely, we obtain the following theorem.

Theorem 3.1 ((uniform) recovery) *There exist absolute constants $E, E', c > 0$ such that the following holds. Let $\epsilon \in (0, 1)$ and assume the GMRA's maximum refinement level $J \geq j := \lceil \log(1/\varepsilon) \rceil$. Further suppose that one has $\operatorname{dist}(\mathbf{0}, \mathcal{M}_j) \geq 1/2$, $0 < C_1 < 2^j$, and $\sup_{\mathbf{x} \in \mathcal{M}} \tilde{C}_\mathbf{x} < 2^{j-2}$. If*

$$m \geq E C_1^{-6} \varepsilon^{-6} \max\left\{ w(\mathcal{M}), \sqrt{d \log(e/\varepsilon)} \right\}^2, \tag{8}$$

then with probability at least $1 - 12e^{-cC_1^2\varepsilon^2 m}$, for all $\mathbf{x} \in \mathcal{M} \subset \mathbb{S}^{D-1}$ the approximations $\mathbf{x}^*$ obtained by OMS satisfy

$$\|\mathbf{x} - \mathbf{x}^*\|_2^2 \leq E'\big(1 + \tilde{C}_\mathbf{x} + C_1 \max\{1, \log C_1^{-1}\}\big)^2 \varepsilon \log \frac{1}{2\varepsilon}. \tag{9}$$

Proof The proof can be found below Theorem 4.14 in Sect. 4. □

Remark 3.2 Let us briefly comment on the assumptions of Theorem 3.1. First, since $\mathcal{M} \subset \mathbb{S}^{D-1}$, requiring $\operatorname{dist}(\mathbf{0}, \mathcal{M}_j) \geq 1/2$ is a mild assumption. Any GMRA not fulfilling it would imply a worst-case reconstruction error of $1/2$ in (9). The constant $1/2$ was chosen for simplicity and can be replaced by an arbitrary number in $(0, 1)$.

This, however, influences the constants E, E', c. Second, the restrictions on C_1 and $\tilde{C}_{\mathbf{x}}$ are easily satisfied, e.g., if the centers form a maximal 2^{-j} packing of $\mathscr{M}$ at each scale j or if the GMRA is constructed from manifold samples as discussed in [38] (cf. Appendix E). In both these cases C_1 and $\tilde{C}_{\mathbf{x}}$ are in fact bounded by absolute constants.

Note that Theorem 3.1 depends on the Gaussian width of $\mathscr{M}$. For general sets K this quantity provides a quite tight measure of the set complexity which might seem counter-intuitive for non-convex K on first sight. After all the convex hull of K might be considerably larger than the set itself while $w(K) = w(\mathrm{conv}\,K)$. However, the intuition is deceptive in this case. In high-dimensional spaces, the intrinsic complexity of non-convex sets and their convex hull hardly differ. For instance, if K is the set of s-sparse vectors in the D-dimensional Euclidean unit ball, its convex hull, the ℓ_1-ball, is full dimensional but $w(K) = w(\mathrm{conv}\,K) \approx s\log(D/s)$ which differs from the information theoretic lower bound on the complexity of K by at most a log-factor.

In the case of compact Riemannian submanifolds of $\mathbb{R}^D$ it might be more convenient to have a dependence on the geometric properties of $\mathscr{M}$ instead (e.g., its volume and reach). Indeed, one can show by means of [17] that $w(\mathscr{M})$ can be upper bounded in terms of the manifold's intrinsic dimension d, its d-dimensional volume $\mathrm{Vol}\,\mathscr{M}$, and the inverse of its reach. Intuitively, these dependencies are to be expected as a manifold with fixed intrinsic dimension d can become more complex as either its volume or curvature (which can be bounded by the inverse of its reach) grows. The following theorem, which is a combination of different results in [17], formalizes this intuition by bounding the Gaussian width of a manifold in terms of its geometric properties.

Theorem 3.3 *Assume $\mathscr{M} \subset \mathbb{R}^D$ is a compact d-dimensional Riemannian manifold with d-dimensional volume $\mathrm{Vol}\,\mathscr{M}$ where $d \geq 1$. Then one can replace $w(\mathscr{M})$ in the above theorem by*

$$w(\mathscr{M}) \leq C \cdot \mathrm{diam}\,\mathscr{M} \cdot \sqrt{d \cdot \max\left\{\log\frac{c\sqrt{d}}{\min\{1, \mathrm{reach}\,\mathscr{M}\}}, 1\right\} + \log\max\{1, \mathrm{Vol}\,\mathscr{M}\}}$$

where $C, c > 0$ are absolute constants.

Proof See Appendix B.   □

Remark 3.4 Note that in our setting $\mathscr{M} \subset \mathbb{S}^{D-1}$ implies that $\mathrm{diam}\,\mathscr{M} \leq 2$ and $\mathrm{reach}\,\mathscr{M} \leq 1$. As we will see, the Gaussian width of the GMRA approximation to $\mathscr{M}$ is also bounded in terms of $w(\mathscr{M})$. This additional width bound is crucial to the proof of Theorem 3.1 as the complexity of the GMRA approximation to $\mathscr{M}$ also matters whenever one attempts to approximate an $\mathbf{x} \in \mathscr{M}$ using only the available GMRA approximation to $\mathscr{M}$. See, e.g., Lemmas 4.3, 4.5, and 4.6 below for upper bounds on the Gaussian widths of GMRA approximations to manifolds $\mathscr{M} \subset \mathbb{S}^{D-1}$ in various settings.

Finally, we point out that Theorem 3.1 assumes access to a GMRA approximation to $\mathscr{M} \subset \mathbb{S}^{D-1}$, which satisfies all of the axioms listed in Definition 2.1. Following the work of Maggioni et al. [38], however, one can also ask whether a similar result will

still hold if the GMRA approximation one has access to has been learned by randomly sampling points from $\mathcal{M}$ without the assumptions of Definition 2.1 being guaranteed a priori. Indeed, such a setting is generally more realistic. In fact it turns out that a version of Theorem 3.1 still holds for such empirical GMRA approximations under suitable conditions; see Theorem E.7. We refer the interested reader to Appendices D and E for additional details and discussion regarding the use of such empirically learned GMRA approximations.

4 Proofs

This section provides proofs of the main result in both settings described above and establishes several technical lemmas. First, properties of the Gaussian width and the geodesic distance are collected and shown. Then, the main results are proven for a given GMRA approximation fulfilling the axioms.

4.1 Toolbox

We start by connecting slightly different definitions of dimensionality measures similar to the Gaussian width and clarify how they relate to each other. This is necessary as the tools we make use of appear in their original versions referring to different definitions of Gaussian width.

Definition 4.1 (Gaussian (mean) width) Let $g \sim \mathcal{N}(\mathbf{0}, \mathrm{Id}_D)$. For a subset $K \subset \mathbb{R}^D$ define

(i) the *Gaussian width*: $w(K) := \mathbb{E}\big[\sup_{\mathbf{x} \in K} \langle \mathbf{g}, \mathbf{x} \rangle\big]$,
(ii) the *Gaussian mean width* to be the Gaussian width of $K - K$, and
(iii) the *Gaussian complexity*: $\gamma(K) = \mathbb{E}\big[\sup_{\mathbf{x} \in K} |\langle \mathbf{g}, \mathbf{x} \rangle|\big]$.

By combining properties 5) and 6) of [41, Prop. 2.1] one has

$$w(K - K) \le 2w(K) \le 2\gamma(K) \le 2\left(w(K - K) + \sqrt{\frac{2}{\pi}} \operatorname{dist}(\mathbf{0}, K)\right). \qquad (10)$$

Remark 4.2 One can easily verify that $w(K) \ge 0$ for all $K \subset \mathbb{R}^D$ since $w(K) := \mathbb{E}\big[\sup_{\mathbf{x} \in K} \langle \mathbf{g}, \mathbf{x} \rangle\big] \ge \sup_{\mathbf{x} \in K} \mathbb{E}[\langle \mathbf{g}, \mathbf{x} \rangle] = 0$. The square $w(K \cap B(0, 1))^2$ of the Gaussian width of $K \subset \mathbb{R}^D$ is also a good measure of intrinsic dimension. For example, if K is a linear subspace with $\dim K = d$ then $w(K \cap \mathcal{B}(\mathbf{0}, 1)) \le \sqrt{d}$. In this sense, the Gaussian width extends the concept of dimension to general sets K. Furthermore, for a finite set K the Gaussian width is bounded by $w(K) \le C_f \operatorname{diam}(K \cup \{\mathbf{0}\})\sqrt{\log |K|}$. This can be deduced directly from the definition (see, e.g., [41, Sect. 2]).

Now that we have introduced the notion of Gaussian width, we can use it to characterize the union of the given manifold and a single level of its GMRA approximation $\mathcal{M} \cup \mathcal{M}_j$ (recall the definition of $\mathcal{M}_j$ in Sect. 2).

Lemma 4.3 (a bound on the Gaussian width for coarse scales) *For $\mathcal{M}_j$, the subspace approximation in the GMRA of level $j > j_0$ (cf. the end of Sect. 2) for $\mathcal{M}$ of dimension $d \geq 1$, the Gaussian width of $\mathcal{M} \cup \mathcal{M}_j$ can be bounded from above and below by*

$$\max\{w(\mathcal{M}), w(\mathcal{M}_j)\} \leq w(\mathcal{M} \cup \mathcal{M}_j)$$
$$\leq 2w(\mathcal{M}) + 2w(\mathcal{M}_j) + 3 \leq 2w(\mathcal{M}) + C\sqrt{dj}.$$

Remark 4.4 Note that the first inequality holds for general sets, not only $\mathcal{M}$ and $\mathcal{M}_j$. Moreover, one only uses $\mathcal{M}_j \subset \mathcal{B}(\mathbf{0}, 2)$ to prove the second inequality. It thus holds for $\mathcal{M}_j$ replaced with arbitrary subsets of $\mathcal{B}(\mathbf{0}, 2)$. We might use both variations referring to Lemma 4.3.

Proof The first inequality follows by noting that

$$\max\{w(\mathcal{M}), w(\mathcal{M}_j)\} = \max\left\{\mathbb{E}\left[\sup_{\mathbf{v}\in\mathcal{M}} \langle \mathbf{v}, \mathbf{g}\rangle\right], \mathbb{E}\left[\sup_{\mathbf{v}\in\mathcal{M}_j} \langle \mathbf{v}, \mathbf{g}\rangle\right]\right\}$$
$$\leq \mathbb{E}\left[\sup_{\mathbf{v}\in\mathcal{M}\cup\mathcal{M}_j} \langle \mathbf{v}, \mathbf{g}\rangle\right] = w(\mathcal{M} \cup \mathcal{M}_j).$$

To obtain the second inequality observe that

$$w(\mathcal{M} \cup \mathcal{M}_j) \leq \gamma(\mathcal{M} \cup \mathcal{M}_j) \leq \mathbb{E}\left[\sup_{\mathbf{v}\in\mathcal{M}} |\langle \mathbf{v}, \mathbf{g}\rangle| + \sup_{\mathbf{v}\in\mathcal{M}_j} |\langle \mathbf{v}, \mathbf{g}\rangle|\right]$$
$$= \gamma(\mathcal{M}) + \gamma(\mathcal{M}_j)$$
$$\leq 2w(\mathcal{M}) + 2w(\mathcal{M}_j) + \sqrt{\frac{2}{\pi}}\,\mathrm{dist}(\mathbf{0}, \mathcal{M}) + \sqrt{\frac{2}{\pi}}\,\mathrm{dist}(\mathbf{0}, \mathcal{M}_j) \qquad (11)$$
$$\leq 2\left(w(\mathcal{M}) + w(\mathcal{M}_j) + 1.5\sqrt{\frac{2}{\pi}}\right),$$

where we used (10), the fact that $\mathcal{M} \subset \mathbb{S}^{D-1}$, and that $\mathcal{M}_j \subset \mathcal{B}(\mathbf{0}, 2)$. For the last inequality we bound $w(\mathcal{M}_j)$. First, note that

$$w(\mathcal{M}_j) = \mathbb{E}\left[\sup_{\mathbf{v}\in\mathcal{M}_j} \langle \mathbf{v}, \mathbf{g}\rangle\right] = \mathbb{E}\left[\sup_{\mathbf{v}\in\{\mathbb{P}_{j,k_j(\mathbf{x})}(\mathbf{x})\,:\,\mathbf{x}\in\mathcal{B}(\mathbf{0},2)\}\cap\mathcal{B}(\mathbf{0},2)} \langle \mathbf{v}, \mathbf{g}\rangle\right]$$
$$\leq \mathbb{E}\left[\sup_{\mathbf{x}\in\bigcup_{k\in[K_j]} P_{j,k}\cap\mathcal{B}(\mathbf{0},2)} \langle \mathbf{x}, \mathbf{g}\rangle\right].$$

For all $k \in [K_j]$ there exist d-dimensional Euclidean balls $L_{j,k} \subset P_{j,k}$ of radius 2 such that $P_{j,k} \cap \mathcal{B}(\mathbf{0}, 2) \subset L_{j,k}$. Hence, $\bigcup_{k\in[K_j]}(P_{j,k} \cap \mathcal{B}(\mathbf{0}, 2)) \subset L_j := \bigcup_{k\in[K_j]} L_{j,k}$. By definition, the ε-covering number of L_j (a union of K_j d-dimensional balls) can be bounded by $\mathcal{N}(L_j, \varepsilon) \leq K_j(6/\varepsilon)^d$ which implies $\log \mathcal{N}(L_j, \varepsilon) \leq dj \log(12C_{\mathscr{C}}/\varepsilon)$

by GMRA property (ii). By Dudley's inequality (see, e.g., [16]) we conclude via Jensen's inequality that

$$
w(\mathscr{M}_j) \leq w(L_j) \leq C_{\text{Dudley}} \int_0^2 \sqrt{\log \mathscr{N}(L_j, \varepsilon)} \, d\varepsilon
$$

$$
\leq C_{\text{Dudley}} \sqrt{dj} \int_0^2 \sqrt{\log 12 C_{\mathscr{C}} - \log \varepsilon} \, d\varepsilon
$$

$$
\leq C_{\text{Dudley}} \sqrt{dj} \sqrt{2 \log 12 C_{\mathscr{C}} - \int_0^2 \log \varepsilon \, d\varepsilon} \leq C' \sqrt{dj},
$$

where C' is a constant depending on C_{Dudley} and $C_{\mathscr{C}}$. Choosing $C = 2C' + 3$ yields the claim as $3\sqrt{2/\pi} \leq 3\sqrt{dj}$.   □

The following two lemmas concerning width bounds for fine scales will also be useful. Their proofs (see Appendix C), though more technical, use ideas similar to the proof of Lemma 4.3. The first lemma improves on Lemma 4.3 for large values of j by considering a more geometrically precise approximation to $\mathscr{M}$, $\mathscr{M}_j^{\text{rel}} \subset \mathscr{M}_j$.

Lemma 4.5 (a bound of the Gaussian width for fine scales) *Assume* $j \geq \log_2 D$, $\max\{1, \sup_{\mathbf{z} \in \mathscr{M}} C_{\mathbf{z}}\} =: C_{\mathscr{M}} < \infty$, *and* $\mathscr{M}_j^{\text{rel}} := \{\mathbb{P}_{j,k_j(\mathbf{z})}(\mathbf{z}) : \mathbf{z} \in \mathscr{M}\} \cap B(\mathbf{0}, 2)$. *We obtain*

$$
\max\{w(\mathscr{M}), w(\mathscr{M}_j^{\text{rel}})\} \leq w(\mathscr{M} \cup \mathscr{M}_j^{\text{rel}})
$$
$$
\leq 2w(\mathscr{M}) + 2w(\mathscr{M}_j^{\text{rel}}) + 3 \leq C(w(\mathscr{M}) + 1) \log D.
$$

It is not surprising that for general $\mathscr{M} \in \mathbb{S}^{D-1}$ the width bound for $w(\mathscr{M}_j)$ (resp. $w(\mathscr{M}_j^{\text{rel}})$) depends on either j or $\log D$. When using the proximity of $\mathscr{M}_j^{\text{rel}}$ to $\mathscr{M}$ in Lemma 4.5 we only use the information that $\mathscr{M}_j^{\text{rel}} \subset \text{tube}_{C_{\mathscr{M}} 2^{-2j}}$ and a large ambient dimension D will lead to a higher complexity of the tube. In the case of Lemma 4.3, we omit the proximity argument by using the maximal number of affine d-dimensional spaces in $\mathscr{M}_j$ and hence do not depend on D but on the refinement level j.

The next lemma just below utilizes even more geometric structure by assuming that $\mathscr{M}$ is a Riemannian manifold. It improves on both Lemma 4.3 and 4.5 for such $\mathscr{M}$ by yielding a width bound which is independent of both j and D for all j sufficiently large.

Lemma 4.6 (a bound of the Gaussian width for approximations to Riemannian manifolds) *Assume* $\mathscr{M} \subset \mathbb{S}^{D-1}$ *is a compact d-dimensional Riemannian manifold with d-dimensional volume* $\text{Vol} \, \mathscr{M}$ *where* $d \geq 1$. *Furthermore, suppose that for* $\max\{1, \sup_{\mathbf{z} \in \mathscr{M}} C_{\mathbf{z}}\} =: C_{\mathscr{M}}$, $j > \max\{j_0, \log_2(8 C_{\mathscr{M}}/C_1)\}$, *and set* $\mathscr{M}_j^{\text{rel}} := \{\mathbb{P}_{j,k_j(\mathbf{z})}(\mathbf{z}) : \mathbf{z} \in \mathscr{M}\} \cap \mathscr{B}(\mathbf{0}, 2)$. *Then there exist absolute constants* $C, c > 0$ *such that*

$$
\max\{w(\mathscr{M}), w(\mathscr{M}_j^{\text{rel}})\} \leq w(\mathscr{M} \cup \mathscr{M}_j^{\text{rel}})
$$

$$\leq C \sqrt{d \left(1 + \log \frac{c\sqrt{d}}{\text{reach } M}\right) + \log \max \{1, \text{Vol } \mathscr{M}\}}.$$

Here the constants C_z and C_1 are from properties (iii.b) *and* (iii.a), *respectively.*

Finally, the following lemma quantifies the equivalence between Euclidean and normalized geodesic distance on the sphere.

Lemma 4.7 *For* $\mathbf{z}, \mathbf{z}' \in \mathbb{S}^{D-1}$ *one has*

$$d_G(\mathbf{z}, \mathbf{z}') \leq \|\mathbf{z} - \mathbf{z}'\|_2 \leq \pi \, d_G(\mathbf{z}, \mathbf{z}').$$

Proof First observe that $\langle \mathbf{z}, \mathbf{z}' \rangle = \cos \angle(\mathbf{z}, \mathbf{z}') = \cos \pi d_G(\mathbf{z}, \mathbf{z}')$. This yields

$$\|\mathbf{z} - \mathbf{z}'\|_2 - d_G(\mathbf{z}, \mathbf{z}') = \sqrt{2 - 2\cos \pi d_G(\mathbf{z}, \mathbf{z}')} - d_G(\mathbf{z}, \mathbf{z}') \geq 0$$

as the function $f(x) = \sqrt{2 - 2\cos \pi x} - x$ is non-negative on $[0, 1]$. For the upper bound note the relation between the geodesic distance $\tilde{d}_G$ and the normalized geodesic distance d_G,

$$\tilde{d}_G(\mathbf{z}, \mathbf{z}') = \pi \, d_G(\mathbf{z}, \mathbf{z}'),$$

which yields

$$\|\mathbf{z} - \mathbf{z}'\|_2 \leq \tilde{d}_G(\mathbf{z}, \mathbf{z}') = \pi \, d_G(\mathbf{z}, \mathbf{z}'). \qquad \square$$

We now have the preliminary results necessary in order to prove Theorem 3.1.

4.2 Proof of Theorem 3.1 with Axiomatic GMRA

Recall that our theoretical result concerns OMS-simple with recovery performed using (3) and (4). The proof is based on the following idea. We first control the error $\|\mathbf{c}_{j,k'} - \mathbf{x}\|_2$ made by (3) in approximating a GMRA center closest to $\mathbf{x}$. To do so we make use of Plan and Vershynin's result on δ-uniform tessellations in [43]. Recall the equivalence between one-bit measurements and random hyperplanes.

Definition 4.8 (uniform tessellation [43, Defn. 1.1]) Let $K \subset \mathbb{S}^{D-1}$ and an arrangement of m hyperplanes in $\mathbb{R}^D$ be given via a matrix A (i.e., the j-th row of A is the normal to the j-th hyperplane). Let $d_A(\mathbf{x}, \mathbf{y}) \in [0, 1]$ denote the fraction of hyperplanes separating $\mathbf{x}$ and $\mathbf{y}$ in K and let d_G be the normalized geodesic distance on the sphere, i.e., opposite poles have distance one. Given $\delta > 0$, the hyperplanes provide a δ-uniform tessellation of K if

$$|d_A(\mathbf{x}, \mathbf{y}) - d_G(\mathbf{x}, \mathbf{y})| \leq \delta$$

holds for all $\mathbf{x}, \mathbf{y} \in K$.

Theorem 4.9 (random uniform tessellation, [43, Thm. 3.1]) *Consider a subset $K \subseteq \mathbb{S}^{D-1}$ and let $\delta > 0$. Let*

$$m \geq \bar{C}\delta^{-6}\max\{w(K)^2, 2/\pi\}$$

and consider an arrangement of m independent random hyperplanes in $\mathbb{R}^D$ uniformly distributed according to the Haar measure. Then with probability at least $1 - 2e^{-c\delta^2 m}$, these hyperplanes provide a δ-uniform tessellation of K. Here and later $\bar{C}, c$ denote positive absolute constants.

Remark 4.10 In words, Theorem 4.9 states that if a number of one-bit measurements scale at least linearly in intrinsic dimension of a set $K \subset \mathbb{S}^{D-1}$, then with high probability the percentage of different measurements of two points x, $y \in K$ is closely related to their distance on the sphere. Implicitly the diameter of all tessellation cells is bounded by δ. The original version of Theorem 4.9 uses $\gamma(K)$ instead of $w(K)$. However, note that by (10) we get for $K \subseteq \mathbb{S}^{D-1}$ that $\gamma(K) \leq w(K - K) + \sqrt{2/\pi} \leq 3w(K)$ as long as $w(K) \geq \sqrt{2/\pi}$ which is reasonable to assume. Hence, if $\bar{C}$ is changed by a factor of 9, Theorem 4.9 can be stated as above.

Using these results we will show in Lemma 4.13 that the center $\mathbf{c}_{j,k'}$ identified in step I of the algorithm OMS-simple satisfies $\|\mathbf{x} - \mathbf{c}_{j,k'}\|_2 \leq 16\max\{\|\mathbf{x} - \mathbf{c}_{j,k_j(\mathbf{x})}\|_2, C_1 2^{-j-1}\}$ in Lemma 4.13. Therefore, the GMRA property (iii.b) provides an upper bound on $\|\mathbf{x} - \mathbb{P}_{j,k'}(\mathbf{x})\|_2$. What remains is to then bound the gap between $\mathbb{P}_{j,k'}(\mathbf{x})$ and the approximation $\mathbf{x}^*$. This happens in two steps. First, Plan and Vershynin's result on noisy one-bit sensing (see Theorem 4.11) is applied to a scaled version of (4) bounding the distance between $\mathbb{P}_{j,k'}(\mathbf{x})$ and $\bar{\mathbf{x}}$ (the minimizer of the scaled version). This argument works by interpreting the true measurements $\mathbf{y}$ as a noisy version of the non-accessible one-bit measurements of $\mathbb{P}_{j,k'}(\mathbf{x})$. The rescaling becomes necessary as Theorem 4.11 is restricted to the unit ball in Euclidean norm. Lastly, a geometric argument is used to bound the distance between the minimum points $\bar{\mathbf{x}}$ and $\mathbf{x}^*$ in order to conclude the proof.

Theorem 4.11 (noisy one-bit [41, Thm. 1.3]) *Let $\mathbf{a}_1, \ldots, \mathbf{a}_m$ be i.i.d. standard Gaussian random vectors in $\mathbb{R}^D$ and let K be a subset of the Euclidean unit ball in $\mathbb{R}^D$. Let $\delta > 0$ and suppose that $m \geq C'\delta^{-6}w(K)^2$. Then with probability at least $1 - 8e^{-c\delta^2 m}$, the following event occurs. Consider a signal $\tilde{\mathbf{x}} \in K$ satisfying $\|\tilde{\mathbf{x}}\|_2 = 1$ and its (unknown) uncorrupted one-bit measurements $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_m)$ given as*

$$\tilde{y}_i = \text{sign}\langle\mathbf{a}_i, \tilde{\mathbf{x}}\rangle, \quad i = 1, 2, \ldots, m.$$

Let $\mathbf{y} = (y_1, \ldots, y_m) \in \{-1, 1\}^m$ be any (corrupted) measurements satisfying $d_{\mathrm{H}}(\tilde{\mathbf{y}}, \mathbf{y}) \leq \tau m$. Then the solution $\bar{\mathbf{x}}$ to the optimization problem

$$\bar{\mathbf{x}} = \arg\max_{\mathbf{z}} \sum_{i=1}^m y_i\langle\mathbf{a}_i, \mathbf{z}\rangle \quad \text{subject to } \mathbf{z} \in K,$$

with input **y***, satisfies*

$$\|\bar{\mathbf{x}} - \tilde{\mathbf{x}}\|_2^2 \leq \delta\sqrt{\log\frac{e}{\delta}} + 11\tau\sqrt{\log\frac{e}{\tau}}.$$

Remark 4.12 Theorem 4.11 yields guaranteed recovery of unknown signals $\mathbf{x} \in K \subset \mathscr{B}(\mathbf{0}, 1)$ up to a certain error by the formulation we use in (4) from one-bit measurements, if the number of measurements scales linearly with the intrinsic dimension of K. The recovery is robust to noise on the measurements. Note that the original version of Theorem 4.11 uses $w(K - K)$ instead of $w(K)$. As $w(K - K) \leq 2w(K)$ by (10), the result stated above also holds for a slightly modified constant C'.

We begin by proving Lemma 4.13.

Lemma 4.13 *If* $m \geq \bar{C}C_1^{-6}2^{6(j+1)}\max\{w(\mathscr{M} \cup \mathbb{P}_{\mathbb{S}}(\mathscr{C}_j))^2, 2/\pi\}$ *the center* $\mathbf{c}_{j,k'}$ *chosen in step I of Algorithm OMS-simple fulfills*

$$\|\mathbf{x} - \mathbf{c}_{j,k'}\|_2 \leq 16\max\{\|\mathbf{x} - \mathbf{c}_{j,k_j(\mathbf{x})}\|_2, C_1 2^{-j-1}\}$$

for all $\mathbf{x} \in \mathscr{M} \subset \mathbb{S}^{D-1}$ *with probability at least* $1 - 2e^{-c(C_1 2^{-j-1})^2 m}$.

Proof By definition of $\mathbf{c}_{j,k'}$ in (3) we have that

$$d_{\mathrm{H}}(\operatorname{sign} A\mathbf{c}_{j,k'}, \mathbf{y}) \leq d_{\mathrm{H}}(\operatorname{sign} A\mathbf{c}_{j,k_j(\mathbf{x})}, \mathbf{y}).$$

As, for all $\mathbf{z}, \mathbf{z}' \in \mathbb{R}^D$, $d_{\mathrm{H}}(\operatorname{sign} A\mathbf{z}, \operatorname{sign} A\mathbf{z}') = m d_A(\mathbf{z}, \mathbf{z}') = m d_A(\mathbb{P}_{\mathbb{S}}(\mathbf{z}), \mathbb{P}_{\mathbb{S}}(\mathbf{z}'))$, this is equivalent to

$$d_A(\mathbb{P}_{\mathbb{S}}(\mathbf{c}_{j,k'}), \mathbf{x}) \leq d_A(\mathbb{P}_{\mathbb{S}}(\mathbf{c}_{j,k_j(\mathbf{x})}), \mathbf{x}).$$

Noting that Gaussian random vectors and Haar random vectors yield identically distributed hyperplanes, Theorem 4.9 now transfers this bound to the normalized geodesic distance, namely

$$d_{\mathrm{G}}(\mathbb{P}_{\mathbb{S}}(\mathbf{c}_{j,k'}), \mathbf{x}) \leq d_{\mathrm{G}}(\mathbb{P}_{\mathbb{S}}(\mathbf{c}_{j,k_j(\mathbf{x})}), \mathbf{x}) + 2\delta$$

with probability at least $1 - 2e^{-c\delta^2 m}$ where $\delta = C_1 2^{-j-1}$. Observe $d_{\mathrm{G}}(\mathbf{z}, \mathbf{z}') \leq \|\mathbf{z} - \mathbf{z}'\|_2 \leq \pi d_{\mathrm{G}}(\mathbf{z}, \mathbf{z}')$ for all $\mathbf{z}, \mathbf{z}' \in \mathbb{S}^{D-1}$ (recall Lemma 4.7) which leads to

$$\|\mathbb{P}_{\mathbb{S}}(\mathbf{c}_{j,k'}) - \mathbf{x}\|_2 \leq \pi d_{\mathrm{G}}(\mathbb{P}_{\mathbb{S}}(\mathbf{c}_{j,k_j(\mathbf{x})}), \mathbf{x}) + 2\pi\delta \leq \pi\|\mathbb{P}_{\mathbb{S}}(\mathbf{c}_{j,k_j(\mathbf{x})}) - \mathbf{x}\|_2 + 2\pi\delta.$$

As by property (iii.a) the centers are close to the manifold, they are also close to the sphere and we have $\|\mathbb{P}_{\mathbb{S}}(\mathbf{c}_{j,k}) - \mathbf{c}_{j,k}\|_2 < C_1 2^{-j-2}$, for all $\mathbf{c}_{j,k} \in \mathscr{C}_j$. Hence, we conclude

$$\|\mathbf{c}_{j,k'} - \mathbf{x}\|_2 \leq \|\mathbf{c}_{j,k'} - \mathbb{P}_{\mathbb{S}}(\mathbf{c}_{j,k'})\|_2 + \|\mathbb{P}_{\mathbb{S}}(\mathbf{c}_{j,k'}) - \mathbf{x}\|_2$$

$$\leq \pi \left(\|\mathbf{c}_{j,k_j(\mathbf{x})} - \mathbf{x}\|_2 + C_1 2^{-j-2} \right) + 2\pi\delta + C_1 2^{-j-2}$$

$$\leq \left(\pi + \frac{\pi}{2} + 2\pi + \frac{1}{2} \right) \max \{ \|\mathbf{c}_{j,k_j(\mathbf{x})} - \mathbf{x}\|_2, C_1 2^{-j-1} \}$$

$$\leq 16 \max \{ \|\mathbf{c}_{j,k_j(\mathbf{x})} - \mathbf{x}\|_2, C_1 2^{-j-1} \}. \qquad \square$$

We can now prove a detailed version of Theorem 3.1 for the given axiomatic GMRA and deduce Theorem 3.1 as a corollary.

Theorem 4.14 (uniform recovery—axiomatic case) *Let $\mathscr{M} \subset \mathbb{S}^{D-1}$ be given by its GMRA for some levels $j_0 < j \leq J$, such that $C_1 < 2^{j_0+1}$ where C_1 is the constant from GMRA properties* (ii.b) *and* (iii.a). *Fix j and assume that $\mathrm{dist}(\mathbf{0}, \mathscr{M}_j) \geq 1/2$. Further, let $d \geq 1$ and*

$$m \geq 16 \max \{ C', \bar{C} \} \cdot C_1^{-6} 2^{6(j+1)} (w(\mathscr{M}) + C\sqrt{dj})^2, \tag{12}$$

where C' is the constant from Theorem 4.11, $\bar{C}$ from Theorem 4.9, and $C > 3$ from Lemma 4.3. Then, with probability at least $1 - 12e^{-(C_1 2^{-j-1})^2 cm}$ the following holds for all $\mathbf{x} \in \mathscr{M}$ with one-bit measurements $\mathbf{y} = \mathrm{sign}\, A\mathbf{x}$ and GMRA constants $\tilde{C}_{\mathbf{x}}$ from property (iii.b) *satisfying $\tilde{C}_{\mathbf{x}} < 2^{j-1}$: The approximations $\mathbf{x}^*$ obtained by OMS fulfill*

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq 2^{-j/2} \sqrt[4]{j}$$
$$\times \left(2\tilde{C}_{\mathbf{x}} 2^{-j/2} + \sqrt{\frac{C_1}{2}} \sqrt[4]{\log \frac{4e}{\min\{C_1, 1\}}} + \sqrt{11 C'_{\mathbf{x}}} \sqrt[4]{\log \frac{2e}{\min\{C'_{\mathbf{x}}, 1\}}} \right).$$

Here $C'_{\mathbf{x}} := 2\tilde{C}_{\mathbf{x}} + C_1$.

Proof of Theorem 3.1 As $j = \lceil \log(1/\varepsilon) \rceil$, we know that $2^{-j} \leq \varepsilon \leq 2^{-j+1}$. This implies

$$m \geq E C_1^{-6} \varepsilon^{-6} \max \left\{ w(\mathscr{M}), \sqrt{d \left(\log \frac{1}{\varepsilon} + 1 \right)} \right\}^2$$
$$\geq 16 \max \{ C', \bar{C} \} \cdot C_1^{-6} 2^{6(j+1)} (w(\mathscr{M}) + C\sqrt{dj})^2$$

for $E > 0$ chosen appropriately. The result follows by applying Theorem 4.14. $\qquad \square$

Proof of Theorem 4.14 Recall that k' is the index chosen by OMS in (6). The proof consists of three steps. First, we apply Lemma 4.13 in **(I)**. By the GMRA axioms this supplies an estimate for $\|\mathbf{x} - \mathbb{P}_{j,k'}(\mathbf{x})\|_2$ with high probability. In **(II)** we use Theorem 4.11 to bound the distance between $\mathbb{P}_{j,k'}(\mathbf{x})/\|\mathbb{P}_{j,k'}(\mathbf{x})\|_2$ and the minimizer $\mathbf{x}^*$ given by

$$x^* = \arg\min_{\mathbf{z}} \sum_{l=1}^{m} (-y_l)\langle \mathbf{a}_l, \mathbf{z}\rangle \quad \text{subject to } \mathbf{z} \in K := \operatorname{conv} \mathbb{P}_{\mathbb{S}}(P_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2))$$

(13)

with high probability. By a union bound over all events, part **(III)** then concludes with an estimate of the distance $\|\mathbf{x} - \mathbf{x}^*\|_2$ combining **(I)** and **(II)**.

(I) Set $\delta := C_1 2^{-j-1}$. Observing that $C_1 2^{-j-2} < 1/2$ by assumption, GMRA property (iii.a) yields that all centers in $\mathscr{C}_j$ are closer to $\mathbb{S}^{D-1}$ than $1/2$, i.e., $1/2 \le \|\mathbf{c}_{j,k}\|_2 \le 3/2$. Hence, by (10)

$$
\begin{aligned}
0 &\le w(\mathbb{P}_{\mathbb{S}}(\mathscr{C}_j)) \le \gamma(\mathbb{P}_{\mathbb{S}}(\mathscr{C}_j)) \\
&\le 2\gamma(\mathscr{C}_j) \le 4w(\mathscr{C}_j) + 2\sqrt{\frac{2}{\pi}}\operatorname{dist}(\mathbf{0}, \mathscr{C}_j) \le 4w(\mathscr{C}_j) + 4.
\end{aligned}
$$

(14)

As $\mathscr{C}_j \subset \mathscr{M}_j$ we know by Lemma 4.3, (14), and Remark 4.4 that

$$
\begin{aligned}
m &\ge 4\bar{C}\delta^{-6}(2w(\mathscr{M}) + 2C\sqrt{dj})^2 \ge 4\bar{C}\delta^{-6}(2w(\mathscr{M}) + 4w(\mathscr{C}_j) + 6)^2 \\
&\ge 4\bar{C}\delta^{-6}(2w(\mathscr{M}) + w(\mathbb{P}_{\mathbb{S}}(\mathscr{C}_j)) + 2)^2 = \bar{C}\delta^{-6}(4w(\mathscr{M}) + 2w(\mathbb{P}_{\mathbb{S}}(\mathscr{C}_j)) + 4)^2 \\
&\ge \bar{C}\delta^{-6}(w(\mathscr{M} \cup \mathbb{P}_{\mathbb{S}}(\mathscr{C}_j)) + 1)^2 \ge \bar{C}\delta^{-6}\max\{w(\mathscr{M} \cup \mathbb{P}_{\mathbb{S}}(\mathscr{C}_j))^2, 2/\pi\}.
\end{aligned}
$$

(15)

Hence, Lemma 4.13 implies that

$$\|\mathbf{x} - \mathbf{c}_{j,k'}\|_2 \le 16 \max\{\|\mathbf{x} - \mathbf{c}_{j,k_j(\mathbf{x})}\|_2, C_1 2^{-j-1}\}$$

with probability at least $1 - 2e^{-c\delta^2 m}$. By GMRA property (iii.b) we now get that

$$\|\mathbf{x} - \mathbb{P}_{j,k'}(\mathbf{x})\|_2 \le \tilde{C}_{\mathbf{x}} 2^{-j}$$

(16)

for some constant $\tilde{C}_{\mathbf{x}}$.

(II) Define $\alpha := \|\mathbb{P}_{j,k'}(\mathbf{x})\|_2$ and note that one has $1/2 \le \alpha \le 3/2$ as $\mathbf{x} \in \mathbb{S}^{D-1}$ and $\|\mathbf{x} - \mathbb{P}_{j,k'}(\mathbf{x})\|_2 \le \tilde{C}_{\mathbf{x}} 2^{-j} \le 1/2$ by (16) and assumption. We now create the setting of Theorem 4.11. Define $\tilde{\mathbf{x}} := \mathbb{P}_{j,k'}(\mathbf{x})/\alpha \in \mathbb{S}^{D-1}$, $\tilde{\mathbf{y}} := \operatorname{sign} A\tilde{\mathbf{x}} = \operatorname{sign} A\mathbb{P}_{j,k'}(\mathbf{x})$, $K = \operatorname{conv}\mathbb{P}_{\mathbb{S}}(P_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2))$, and $\tau := (2\tilde{C}_{\mathbf{x}} + C_1)2^{-j}$. If successfully applied with these quantities, Theorem 4.11 will bound $\|\tilde{\mathbf{x}} - \mathbf{x}^*\|_2$ by

$$
\begin{aligned}
\|\tilde{\mathbf{x}} - \mathbf{x}^*\|_2 &\le \sqrt{\delta\sqrt{\log\frac{e}{\delta}} + 11\tau\sqrt{\log\frac{e}{\tau}}} \\
&\le \left(\sqrt{\frac{C_1}{2}}\sqrt[4]{\log\frac{4e}{\min\{C_1, 1\}}}\right. \\
&\quad \left. + \sqrt{11(2\tilde{C}_x + C_1)}\sqrt[4]{\log\frac{2e}{\min\{2\tilde{C}_x + C_1, 1\}}}\right)2^{-j/2}\sqrt[4]{j}.
\end{aligned}
$$

(17)

All that remains is to verify that the conditions of Theorem 4.11 are met so that (17) is guaranteed with high probability. We first have to check $d_H(\tilde{\mathbf{y}}, \mathbf{y}) \leq \tau m$. Recall that $1/\alpha \leq 2$ and for $\alpha > 0$ one has $\alpha w(K) = w(\alpha K)$. Applying Lemma 4.3 and (10) we have, in analogy to (15), that

$$
\begin{aligned}
m &\geq \bar{C}\delta^{-6}(4w(\mathcal{M}) + 4w(\mathcal{M}_j) + 12)^2 \geq \bar{C}\delta^{-6}\left(2w(\mathcal{M}) + 2w\left(\frac{\mathcal{M}_j}{\alpha}\right) + 12\right)^2 \\
&\geq \bar{C}\delta^{-6}\left(w\left(\mathcal{M} \cup \frac{\mathcal{M}_j}{\alpha}\right) + 7\right)^2 \geq \bar{C}\delta^{-6}\left(w\left(\left(\mathcal{M} \cup \frac{\mathcal{M}_j}{\alpha}\right) \cap \mathscr{B}(\mathbf{0}, 1)\right) + 7\right)^2.
\end{aligned}
$$

Note that in the third inequality a slight modification of the second inequality in Lemma 4.3 is used. As $\mathcal{M}_j/\alpha \subset \mathscr{B}(\mathbf{0}, 4)$, one has $w(\mathcal{M} \cup \mathcal{M}_j/\alpha) \leq 2w(\mathcal{M}) + 2w(\mathcal{M}_j/\alpha) + 5$ by adapting (11). We can now use Theorem 4.9, Lemma 4.7, and the fact that $|1 - \alpha| = |\|\mathbf{x}\|_2 - \|\mathbb{P}_{j,k'}(\mathbf{x})\|_2| \leq \|\mathbf{x} - \mathbb{P}_{j,k'}(\mathbf{x})\|_2$ to obtain

$$
\begin{aligned}
\frac{d_H(\tilde{\mathbf{y}}, \mathbf{y})}{m} &= d_A(\tilde{\mathbf{x}}, \mathbf{x}) \leq d_G(\tilde{\mathbf{x}}, \mathbf{x}) + \delta \leq \|\tilde{\mathbf{x}} - \mathbf{x}\|_2 + \delta \\
&\leq \|\tilde{\mathbf{x}} - \mathbb{P}_{j,k'}(\mathbf{x})\|_2 + \|\mathbb{P}_{j,k'}(\mathbf{x}) - \mathbf{x}\|_2 + \delta \\
&= |1 - \alpha| + \|\mathbb{P}_{j,k'}(\mathbf{x}) - \mathbf{x}\|_2 + \delta \\
&\leq 2\|\mathbb{P}_{j,k'}(\mathbf{x}) - \mathbf{x}\|_2 + \delta \leq (2\tilde{C}_{\mathbf{x}} + C_1)2^{-j} = \tau
\end{aligned}
$$

with probability at least $1 - 2e^{-c\delta^2 m}$. Furthermore, by a similar argumentation as in (14) one gets

$$
w(K) = w(\mathbb{P}_{\mathbb{S}}(P_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2))) \leq 4w(\mathcal{M}_j) + 4, \tag{18}
$$

where one uses invariance of the Gaussian width under taking the convex hull (see [41, Proposition 2.1]), the fact that $P_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2) \subset \mathcal{M}_j$, and the assumption that $1/2 \leq \mathrm{dist}(\mathcal{M}_j, \mathbf{0}) \leq 2$. In combination with Lemma 4.3 we have, in analogy to (15), that

$$
m \geq 4C'\delta^{-6}(2w(\mathcal{M}) + 4w(\mathcal{M}_j) + 6)^2 \geq 4C'\delta^{-6}(w(K) + 2)^2 \geq C'\delta^{-6}w(K)^2.
$$

Hence, we can apply Theorem 4.11 to obtain with probability at least $1 - 8e^{-c\delta^2 m}$ that

$$
\|\bar{\mathbf{x}} - \tilde{\mathbf{x}}\|_2^2 \leq \delta\sqrt{\log\frac{e}{\delta}} + 11\tau\sqrt{\log\frac{e}{\tau}};
$$

the estimate (17) now follows.

(III) To conclude the proof we apply a union bound and obtain with probability at least $1 - 12e^{-c\delta^2 m}$ that

$$\begin{aligned}
\|\mathbf{x} - \mathbf{x}^*\|_2 &\leq \|\mathbf{x} - \mathbb{P}_{j,k'}(\mathbf{x})\|_2 + \|\mathbb{P}_{j,k'}(\mathbf{x}) - \tilde{\mathbf{x}}\|_2 + \|\tilde{\mathbf{x}} - \mathbf{x}^*\|_2 \\
&= \|\mathbf{x} - \mathbb{P}_{j,k'}(\mathbf{x})\|_2 + |1 - \alpha| + \|\tilde{\mathbf{x}} - \mathbf{x}^*\|_2 \\
&\leq 2\|\mathbf{x} - \mathbb{P}_{j,k'}(\mathbf{x})\|_2 + \|\tilde{\mathbf{x}} - \mathbf{x}^*\|_2.
\end{aligned}$$

GMRA property (iii.b) combined with (17) now yields the final desired error bound.
$\square$

Remark 4.15 For obtaining the lower bounds on m in (12) and (8) we made use of Lemma 4.3 leading to the influence of j which is suboptimal for fine scales (i.e., j large). To improve on this for large j one can exploit the alternative versions of the lemma, namely, Lemmas 4.5 and 4.6. Then, however, some minor modifications become necessary in the proof of Theorem 4.14 as the lemmas only apply to $\mathcal{M}_j^{\mathrm{rel}}$:

– In **(I)**, e.g., one has to guarantee that $\mathscr{C}_j \subset \mathcal{M}_j^{\mathrm{rel}}$, i.e., that each center $\mathbf{c}_{j,k}$ is a best approximation for some part of the manifold. This is a reasonable assumption especially if the centers are constructed as means of small manifold patches, which is a common approach in empirical applications (cf. Appendix D).
– Also, when working with $\mathcal{M}_j^{\mathrm{rel}}$ it is essential in **(II)** to have a near-best approximation subspace of $\mathbf{x}$, i.e., the k' obtained in **(I)** has to fulfill $k' \approx k_j(\mathbf{x})$ as $\mathcal{M}_j^{\mathrm{rel}}$ does not include many near-optimal centers for each point on $\mathcal{M}$. Here, one can exploit the minimal distance of centers $\mathbf{c}_{j,k}$ to each other as described in GMRA property (ii.b) and choose δ slightly smaller (in combination with a correspondingly strengthened upper bound in Lemma 4.13) to obtain the necessary guarantees for **(I)**. As we are principally concerned with the case where $j = \mathcal{O}(\log D)$ in this paper, however, we will leave such variants to future work.

We are now prepared to explore the numerical performance of the proposed methods.

5 Numerical Simulation

In this section, we present various numerical experiments to benchmark OMS. The GMRAs we work with are constructed using the GMRA code provided by Maggioni.[2] We compared the performance of OMS for three exemplary choices of $\mathcal{M}$, namely, a simple 2-dim sphere embedded in $\mathbb{R}^{20}$ (20000 data points sampled from the 2-dimensional sphere $\mathcal{M}$ embedded in $\mathbb{S}^{20-1}$), the MNIST data set [36] of handwritten digits "1" (3000 data points in $\mathbb{R}^{784}$), and the Fashion-MNIST data set [47] of shirt images (2000 data points in $\mathbb{R}^{784}$). Both MNIST data sets have been projected to the unit sphere before taking measurements and computing the GMRA. In each of the experiments 5.1–5.4, we first computed a GMRA up to refinement level $j_{\max} = 10$ and then recovered 100 randomly chosen $\mathbf{x} \in \mathcal{M}$ from their one-bit measurements

---

[2] The code is available at http://www.math.jhu.edu/~mauro/#tab_code.

by applying OMS. Depicted is the averaged relative error between $\mathbf{x}$ and its approximation $\mathbf{x}^*$, i.e., $\|\mathbf{x} - \mathbf{x}^*\|_2 / \|\mathbf{x}\|_2$ which is equal to the absolute error $\|\mathbf{x} - \mathbf{x}^*\|_2$ for $\mathscr{M} \subset \mathbb{S}^{D-1}$. Note the different approximation error ranges of the sphere and the MNIST experiments when comparing both settings. As a benchmark, the average error caused by the best GMRA approximation, i.e., projection of $\mathbf{x}$ onto the GMRA, is provided in all plots in the form of a horizontal dashed black line.[3] Let us mention that the error caused by best GMRA approximation—though being a benchmark for the recovery performance one could expect—is not a strict lower bound to OMS since the algorithm is not restricted to the GMRA but uses convex relaxations of the GMRA subspaces, cf. Fig. 4a.

5.1 OMS-Simple vs. OMS

The first test compares recovery performance of the two algorithms presented above, namely OMS-simple for $R \in \{0.5, 1, 1.5\}$ and OMS. The results are depicted in Fig. 4. Note that only $R = 1.5$ and, in the case of the 2-sphere, $R = 1$ are depicted as in the respective other cases for each number of measurements most of the trials did not yield a feasible solution in (4), so the average was not well defined. One can observe that for all data sets OMS outperforms OMS-simple which is not surprising as OMS does not rely on a suitable parameter choice. This observation is also the reason for us to restrict the theoretical analysis to OMS. The more detailed approximation of the toy example (2-dimensional sphere) is due to its simpler structure and lower dimensional setting and can also be observed in 5.2–5.4. Figures 5 and 6 depict one specific reconstructed image of Fashion-MNIST, for four different numbers of measurements m. Obviously, OMS shows a better performance having less quantization artifacts. Moreover, the good visual quality of the OMS reconstruction in Fig. 6 for only $m = 100$ suggests that the ℓ_2-error used in Fig. 4 is a rather pessimistic performance measure. Considering that all MNIST-images could be fully coded in $8 \cdot 28 \cdot 28 = 6272$ bits (gray-scale images contain only 8-bit of information per pixel), it is important to point out the potentially overly pessimistic nature of the ℓ_2-errors reported for $m = 10000$, as well as to note that the visual quality is already much better than one might expect for OMS at compressions of ratios of size $100/6272 < 0.02$.

5.2 Modifying OMS

Observations in [34] motivate to consider a modification of OMS in which (7) is replaced by

$$\mathbf{x}^* = \underset{\mathbf{z} \in \mathbb{R}^D}{\arg\min} \sum_{l=1}^m [(-y_l)\langle \mathbf{a}_l, \mathbf{z} \rangle]_+ \quad \text{subject to} \quad \mathbf{z} \in \operatorname{conv} \mathbb{P}_{\mathbb{S}}(P_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2)),$$

---

[3] To be precise, for each $\mathbf{x}$ we picked the GMRA subspace minimizing the projection distance $\|\mathbf{x} - \mathbb{P}_{j,k}(\mathbf{x})\|_2$ and averaged this error over all realizations of $\mathbf{x}$.

(a) 2-sphere

(b) MNIST

(c) Fashion-MNIST
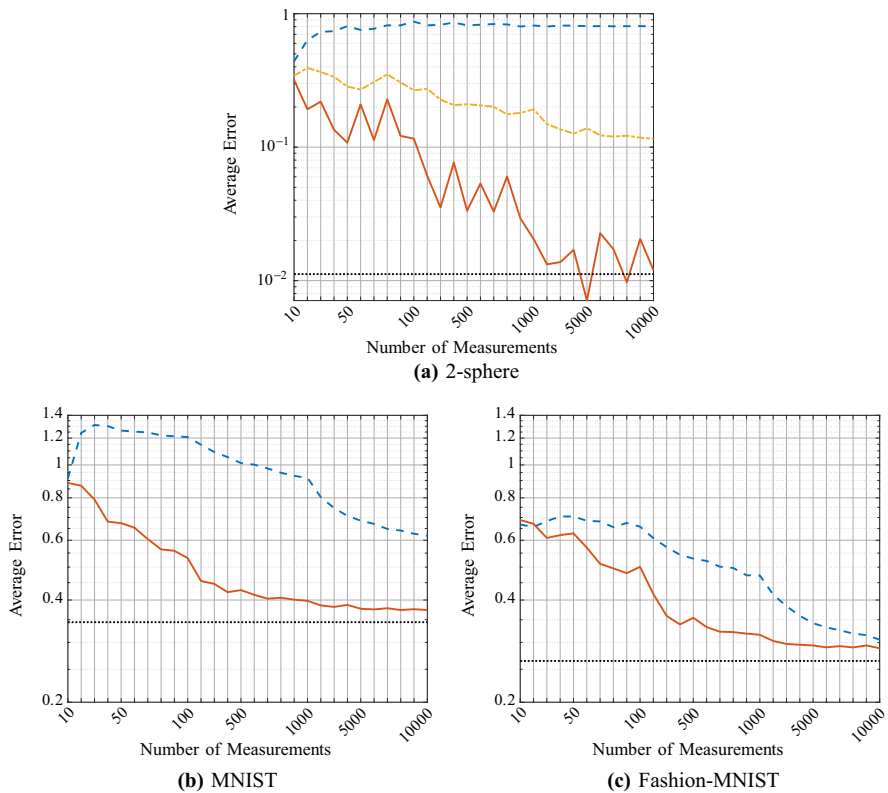
Fig. 4 Comparison of OMS-Simple for $R = 1$ (dotted-dashed, yellow), $R = 1.5$ (dashed, blue), and OMS (solid, red). Recall from Sect. 5.1 that OMS-Simple for $R = 1$ does not recover in **b** and **c**. The black dotted line highlights average error caused by direct GMRA projection

(a) Groundtruth   **(b)** GMRA   **(c)** $m=10$ **(d)** $m=100$ **(e)** $m=1000$ **(f)** $m=10000$
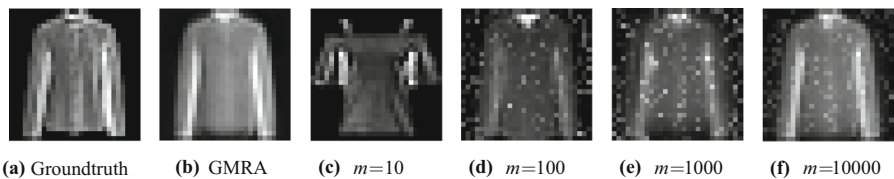
Fig. 5 One Fashion-MNIST signal with its OMS-Simple ($R = 1.5$) reconstructions. The best GMRA approximation is given as a benchmark. Note that the GMRA uses an 11-dimensional subspace at this part of the manifold

(a) Groundtruth   **(b)** GMRA   **(c)** $m=10$ **(d)** $m=100$ **(e)** $m=1000$ **(f)** $m=10000$
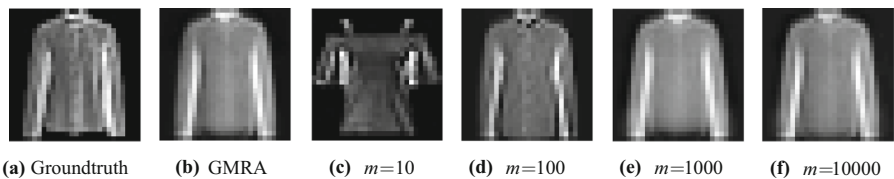
Fig. 6 One Fashion-MNIST signal with its OMS reconstructions. The best GMRA approximation is given as a benchmark. Note that the GMRA uses an 11-dimensional subspace at this part of the manifold

where $[t]_+ = \max\{0, t\}$ denotes the positive part of $t \in \mathbb{R}$. We tested this approach in some initial numerical experiments, but found that the modification produced ambiguous results with no clear improvement. Let us mention that after the original version of this paper had been finished, the $[\,\cdot\,]_+$-formulation has been thoroughly analyzed for robust one- and multi-bit quantization in a dithered measurement setup in [32], but not in the context of GMRA and with dithering (this explains the ambiguous experimental outcomes in our case, as we do not use dithering in the one-bit measurements). Certainly one could transfer results from [32] to our setting by changing the measurement model to include dithering, cf. [13].

5.3 Are Two Steps Necessary?

One might wonder if the two steps in OMS-simple and OMS are necessary at all. Wouldn't it be sufficient to use the center $\mathbf{c}_{j,k'}$ determined in step I as an approximation for $\mathbf{x}$? If the GMRA is fine enough, this indeed is the case. If one only has access to a rather rough GMRA, the simulations in Fig. 7 show that the second step makes a notable difference in approximation quality. This behavior is not surprising in view of Lemma 4.13. The lemma guarantees a good approximation of $\mathbf{x}$ by $\mathbf{c}_{j,k'}$ as long as $\mathbf{x}$ is well approximated by an optimal center. For both MNIST data sets, one can observe that the second step only improves performance if the number of one-bit measurements is sufficiently high. For a small set of measurements the centers might yield better approximation as they lie close to $\mathscr{M}$ by GMRA property (iii.a). On the other hand, only parts of the affine spaces are practical for approximation and a certain number of measurements is necessary to restrict II to the relevant parts.
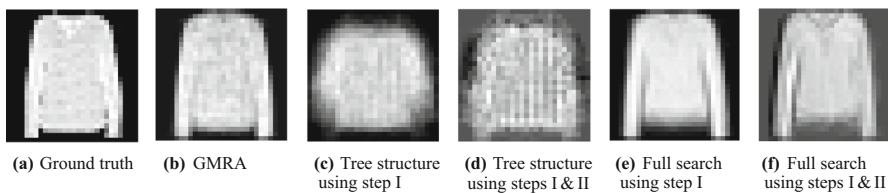
5.4 Tree vs. No Tree

In the fourth test, we checked if approximation still works when not all possible centers are compared in step I of OMS but their tree structure is used. More precisely, to find an optimal center one compares on the first refinement level all centers, and then continues in each subsequent level solely with the children of the k best centers (in the presented experiments we chose $k = 10$). Of course, the chosen center will not be optimal as not all centers are compared (see Fig. 7). In the simple 2-dimensional sphere setting, step II, however, can compensate the worse approximation quality of I with tree search. Figure 7a hardly shows a difference in final approximation quality in both cases. In both MNIST settings, however, one can observe a considerable difference even when performing two steps. Figures 8 and 9 illustrate the differences in reconstruction using tree search vs. full center comparison, respectively, using only step I of OMS vs. using both steps, for $m = 100$ and $m = 1000$. This corresponds to a compression ratio of $100/6272 < 0.02$ resp. $1000/6272 < 0.2$, cf. the discussion in Sect. 5.1. Comparing c and d with e and f of Fig. 8 one sees that full center search reconstructs more detailed features (shape), comparing c and d with e and f of Fig. 9 one sees that the second step enhances reconstruction of details (neckline).

(a) 2-Sphere

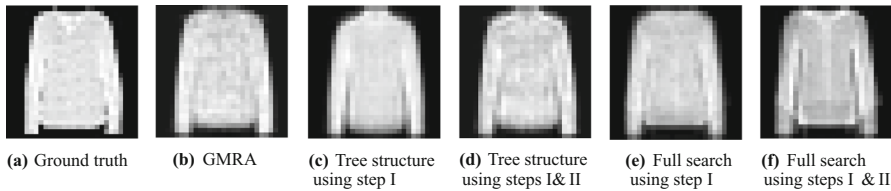**(b)** MNIST

(c) Fashion-MNIST
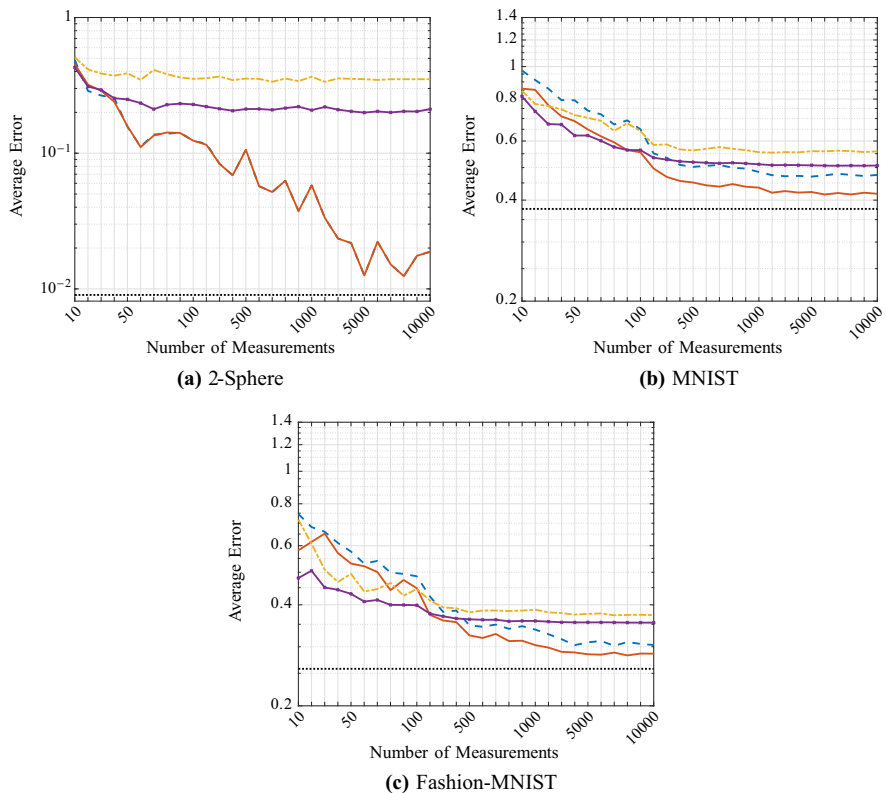
Fig. 7 Comparison of the following: Approximation by step I of OMS when using tree structure (dashed with points, yellow) and when comparing all centers (solid with points, purple); approximation by steps I & II of OMS when using tree structure (dashed, blue; this line is mostly hidden behind the solid red curve in the first plot) and when comparing all centers (solid, red). The black dotted line highlights average error caused by direct GMRA projection

(a) Ground truth **(b)** GMRA **(c)** Tree structure using step I **(d)** Tree structure using steps I & II **(e)** Full search using step I **(f)** Full search using steps I & II

Fig. 8 One Fashion-MNIST signal with OMS reconstructions using tree search or full center search and using only step I or both steps, for $m = 100$. The best GMRA approximation is given as a benchmark. Note that the GMRA uses a 7-dimensional subspace at this part of the manifold

(a) Ground truth   **(b)** GMRA   **(c)** Tree structure   **(d)** Tree structure   **(e)** Full search   **(f)** Full search
                                              using step I   using steps I& II   using step I   using steps I & II

Fig. 9 One Fashion-MNIST signal with OMS reconstructions using tree search or full center search and using only step I or both steps, for $m = 1000$. The best GMRA approximation is given as a benchmark. Note that the GMRA uses a 7-dimensional subspace at this part of the manifold

5.5 A Change of Refinement Level

The last experiment (see Fig. 10) examines the influence of the refinement level j on the approximation error. For small j (corresponding to a rough GMRA) a high number of measurements can hardly improve the approximation quality while for large j (corresponding to a fine GMRA) the approximation error decreases with increasing measurement rates. This behavior is as expected. A rough GMRA cannot profit much from many measurements as the GMRA approximation itself yields an approximate lower bound on the obtainable approximation error. For fine GMRAs the behavior along the measurement axis is similar to above experiments. Note that further increase of j for the same range of measurements did not improve accuracy. Notably, the Fashion-MNIST reconstruction performs quite well even for small choices of j suggesting that the manifold of shirt images is, at least in terms of GMRA approximation, of lower complexity than the manifold of digits "1." This is in line with the observation that the approximation error for Fashion-MNIST is smaller than for MNIST in the above experiments. The non-monotonous increase of the quality of reconstruction in Fig. 10c—the levels $j = 4, 8, 10$ perform especially well—is more difficult to explain; it might relate to the fact that Fashion-MNIST is considered to be harder to classify than MNIST—there seem to be features identifying the images (overall shape, detailed contours) on various levels of refinement.

6 Discussion

In this paper, we proposed OMS, a tractable algorithm to approximate data lying on low-dimensional manifolds from compressive one-bit measurements, thereby complementing the theoretical results of Plan and Vershynin on one-bit sensing for general sets in [41] in this important setting. We then proved (uniform) worst-case error bounds for approximations computed by OMS under slightly stronger assumptions than [41], and also performed numerical experiments on both toy examples and real-world data. As a byproduct of our theoretical analysis (see, e.g., Sect. 4), we have further linked the theoretical understanding of one-bit measurements as tessellations of the sphere [43] to the GMRA techniques introduced in [3], by analyzing the interplay between a given manifold and its GMRA approximation's complexity measured in terms of the Gaussian mean width. Finally, to indicate applicability of our results we showed that

(a) 2-Sphere

(b) MNIST

(c) Fashion-MNIST
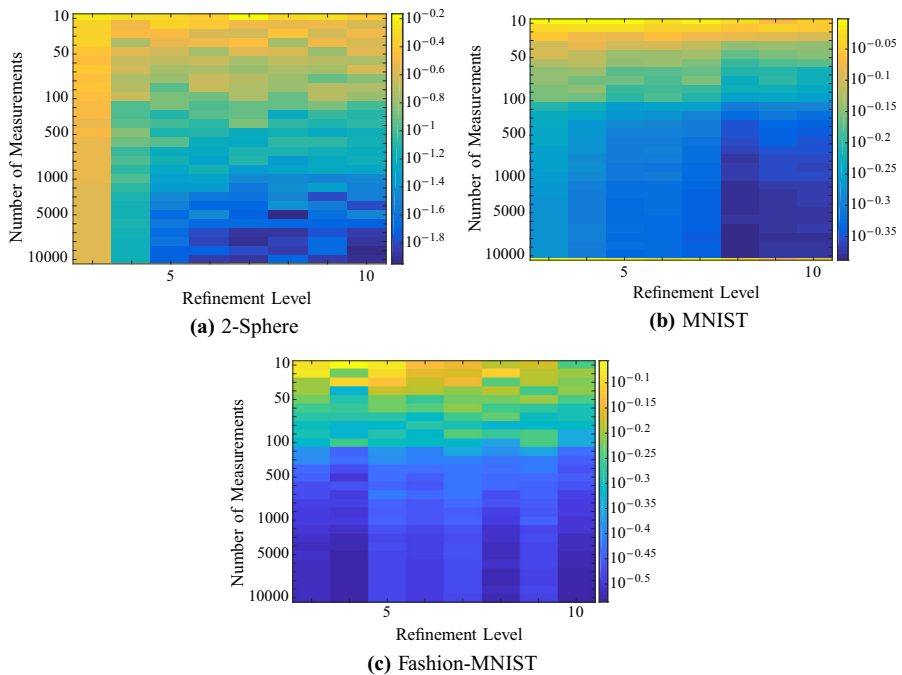
Fig. 10 Approximation error of OMS for different refinement levels j and numbers of measurements

they hold even if there are just random samples from the manifold at hand as opposed to the entire manifold (see, e.g., Appendices D and E). Several interesting questions remain for future research, however.

First, the experiments in Sect. 5.4 suggest a possible benefit from using the tree structure within $\mathscr{C}_j$. Indeed, approximation of OMS does still yield comparable results if I is restricted to a tree-based search which has the advantage of being computable much faster than the minimization over all possible centers. It would be desirable to obtain theoretical error bounds even in this case, as well as to consider the use of other related fast nearest neighbor methods from computer science [26].

Second, the attentive reader might have noticed in the empirical setting of Appendices D and E that (**A2**) in combination with Lemma E.6 seems to imply that II of OMS may be unnecessary. As can be seen from Sect. 5.3 though, the second step of OMS yields a notable improvement even with an empirically constructed GMRA which hints that even with (**A2**) not strictly fulfilled the empirical GMRA techniques remain valid, and II of OMS of value. Understanding this phenomenon might lead to more relaxed assumptions than (**A1**)–(**A4**).

Third, it could be rewarding to also consider versions of OMS for additional empirical GMRA variants including, e.g., those which rely on adaptive constructions [37], GMRA constructions in which subspaces that minimize different criteria are used to approximate the data in each partition element (see, e.g., [27]), and distributed GMRA constructions which are built up across networks using distributed clustering [4] and

SVD [30] algorithms. Such variants could prove valuable with respect to reducing the overall computational storage and/or runtime requirements of OMS in different practical situations.

Fourth, it would be illustrative to compare our GMRA approach which can be interpreted as "signal processing"-type convexification (or at least reduction to a convex subproblem) with recent advances on non-convex ADMM [35], i.e., direct application of non-convex optimization techniques. Two approaches seem of special interest: first, one could replace the GMRA manifold model by Generative Adversarial Networks (GANs) and apply ADMM to the resulting non-convex reconstruction problem or, second, stay in the GMRA setting but not restrict the reconstruction procedure to one single subspace.

Finally, it would also likely be fruitful to explore alternate quantization schemes to the one-bit measurements (1) considered herein. In particular, the so-called $\Sigma\Delta$ quantization schemes generally outperform the type of memoryless scalar quantization methods considered herein both theoretically and empirically in compressive sensing contexts [25,33,45], and initial work suggests that they may provide similar improvements in the GMRA model setting considered here [28]. Nevertheless, one should be aware that feedback quantization schemes like $\Sigma\Delta$ quantization might be of limited practicability in recent large-scale applications like massive MIMO [24], where the measurements are collected (and quantized) in a distributed fashion rendering the use of feedback information impossible. In such cases, the just analyzed memoryless quantization model is of particular interest.

Appendix A: Characterization of Convex Hull

Lemma A.1 *Let $P_{j,k'}$ be the affine subspace chosen in step I of OMS-simple and define* $\mathbf{c} = \mathbb{P}_{j,k'}(\mathbf{0})$. *If* $\mathbf{0} \notin P_{j,k'}$, *the following equivalence holds:*

$$\mathbf{z} \in \operatorname{conv} \mathbb{P}_{\mathbb{S}}(P_{j,k'} \cap \mathcal{B}(\mathbf{0}, 2)) \quad \Leftrightarrow \quad \begin{cases} \|\mathbf{z}\|_2 \leq 1, \\ \Phi_{j,k'}^T \Phi_{j,k'} \mathbf{z} + \mathbb{P}_{\mathbf{c}}(\mathbf{z}) = \mathbf{z}, \\ \langle \mathbf{z}, \mathbf{c} \rangle \geq \|\mathbf{c}\|_2^2/2. \end{cases} \quad (19)$$

Proof First, assume $\mathbf{z} \in \operatorname{conv} \mathbb{P}_{\mathbb{S}}(P_{j,k'} \cap \mathcal{B}(\mathbf{0}, 2))$. Obviously, $\|\mathbf{z}\|_2 \leq 1$. As projecting onto the sphere is a simple rescaling, $\operatorname{conv} \mathbb{P}_{\mathbb{S}}(P_{j,k'} \cap \mathcal{B}(\mathbf{0}, 2)) \subset \operatorname{span} P_{j,k'}$ implies that $\Phi_{j,k'}^T \Phi_{j,k'} \mathbf{z} + \mathbb{P}_{\mathbf{c}}(\mathbf{z}) = \mathbf{z}$. For showing the third constraint note that any $\mathbf{z}' \in P_{j,k'}$ can be written as $\mathbf{z}' = \mathbf{c} + (\mathbf{z}' - \mathbf{c})$ where $\mathbf{z}' - \mathbf{c}$ is perpendicular to $\mathbf{c}$. If in addition $\|\mathbf{z}'\|_2 \leq 2$, we get

$$\langle \mathbb{P}_{\mathbb{S}}(\mathbf{z}'), \mathbf{c} \rangle = \left\langle \frac{\mathbf{z}'}{\|\mathbf{z}'\|_2}, \mathbf{c} \right\rangle = \frac{\langle \mathbf{c}, \mathbf{c} \rangle}{\|\mathbf{z}'\|_2} \geq \frac{\|\mathbf{c}\|_2^2}{2}.$$

As $\mathbf{z}$ is a convex combination of different $\mathbb{P}_{\mathbb{S}}(\mathbf{z}')$, the constraint also holds for $\mathbf{z}$. Let $\mathbf{z}$ fulfill the three constraints. Then $\mathbf{z}' = (\|\mathbf{c}\|_2^2/\langle \mathbf{z}, \mathbf{c} \rangle) \cdot \mathbf{z}$ satisfies $\mathbf{z}' \in P_{j,k'}$ because of the second constraint and $\langle \mathbf{z}', \mathbf{c} \rangle = \|\mathbf{c}\|_2^2$. Furthermore, by the first and third constraints, $\|\mathbf{z}'\|_2 \leq (\|\mathbf{c}\|_2^2/\langle \mathbf{z}, \mathbf{c} \rangle) \leq 2$ and hence $\mathbf{z}' \in P_{j,k'} \cap \mathcal{B}(\mathbf{0}, \|\mathbf{c}\|_2^2/\langle \mathbf{z}, \mathbf{c} \rangle) \subset P_{j,k'} \cap \mathcal{B}(\mathbf{0}, 2)$. As $P_{j,k'} \cap \mathcal{B}(\mathbf{0}, \|\mathbf{c}\|_2^2/\langle \mathbf{z}, \mathbf{c} \rangle)$ is the convex hull of $P_{j,k'} \cap (\|\mathbf{c}\|_2^2/\langle \mathbf{z}, \mathbf{c} \rangle) \cdot \mathbb{S}^{D-1}$, there are $\mathbf{z}_1, \ldots, \mathbf{z}_n \in P_{j,k'}$ and $\lambda_1, \ldots, \lambda_n \geq 0$ with $\|\mathbf{z}_k\|_2 = \|\mathbf{c}\|_2^2/\langle \mathbf{z}, \mathbf{c} \rangle$ and $\sum \lambda_k = 1$ such that $(\|\mathbf{c}\|_2^2/\langle \mathbf{z}, \mathbf{c} \rangle) \cdot \mathbf{z} = \sum \lambda_k \mathbf{z}_k$. Hence, $\mathbf{z} = \sum \lambda_k (\langle \mathbf{z}, \mathbf{c} \rangle/\|\mathbf{c}\|_2^2) \cdot \mathbf{z}_k$. As $(\langle \mathbf{z}, \mathbf{c} \rangle/\|\mathbf{c}\|_2^2) \cdot \mathbf{z}_k \in \mathbb{P}_{\mathbb{S}}(P_{j,k'} \cap \mathcal{B}(\mathbf{0}, 2))$ we get $\mathbf{z} \in \operatorname{conv} \mathbb{P}_{\mathbb{S}}(P_{j,k'} \cap \mathcal{B}(\mathbf{0}, 2))$. $\square$

Appendix B: Proof of Theorem 3.3

Denote by τ the reach of $\mathcal{M}$ and by ρ the diameter $\operatorname{diam} \mathcal{M}$. First, note that for a set $K \subset \mathbb{R}^D$, by Dudley's inequality [16],

$$w(K) \leq C' \int_0^{\operatorname{diam} K/2} \sqrt{\log \mathcal{N}(K, \varepsilon)} \, d\varepsilon,$$

where C' is an absolute constant. Second, [17, Lem. 14] states that the covering number $\mathcal{N}(\mathcal{M}, \varepsilon)$ of a d-dimensional Riemannian manifold $\mathcal{M}$ can be bounded by

$$\mathcal{N}(\mathcal{M}, \varepsilon) \leq \left(\frac{2}{\varepsilon \sqrt{1 - (\varepsilon/(4\tau))^2}} \right)^d \frac{\operatorname{Vol} \mathcal{M}}{\operatorname{Vol} \mathcal{B}_d}$$

for $\varepsilon \leq \tau/2$. After noting that $\operatorname{Vol} \mathcal{B}_d \geq \beta^{-1}(2\pi/d)^{d/2}$ for all $d \geq 1$ for an absolute constant $\beta > 1$, this expression may be simplified to

$$\mathcal{N}(\mathcal{M}, \varepsilon) \leq \beta \left(\frac{\sqrt{2d}}{\varepsilon \sqrt{\pi} \sqrt{1 - (\varepsilon/(4\tau))^2}} \right)^d \operatorname{Vol} \mathcal{M} \leq \beta \left(\frac{\sqrt{d}}{\varepsilon \sqrt{1 - (\varepsilon/(4\tau))^2}} \right)^d \operatorname{Vol} \mathcal{M}.$$

We can combine these facts to obtain

$$
w(\mathcal{M}) \le C' \int_0^{\rho/2} \sqrt{\log \mathcal{N}(\mathcal{M}, \varepsilon)}\, d\varepsilon \le C' \left(\frac{\rho}{2} \int_0^{\rho/2} \log \mathcal{N}(\mathcal{M}, \varepsilon)\, d\varepsilon \right)^{1/2}
$$

$$
= C' \sqrt{\frac{\rho}{2}} \left(\int_0^{\tau/2} \log \mathcal{N}(\mathcal{M}, \varepsilon)\, d\varepsilon + \int_{\tau/2}^{\rho/2} \log \mathcal{N}(\mathcal{M}, \varepsilon)\, d\varepsilon \right)^{1/2},
$$

by using Cauchy–Schwarz inequality for the second inequality. We now bound the first integral by

$$
\int_0^{\tau/2} \log \mathcal{N}(\mathcal{M}, \varepsilon)\, d\varepsilon \le \int_0^{\tau/2} \left[-d \log\left(\frac{\varepsilon}{\beta\sqrt{d}} \underbrace{\sqrt{1 - \frac{\varepsilon^2}{16\tau^2}}}_{\ge 1/2,\text{ as } \varepsilon \le \tau/2} \right) + \log(\mathrm{Vol}\,\mathcal{M}) \right] d\varepsilon
$$

$$
\le \int_0^{\tau/2} -d \log \frac{\varepsilon}{2\beta\sqrt{d}}\, d\varepsilon + \frac{\tau}{2} \log(\mathrm{Vol}\,\mathcal{M})
$$

$$
= -d \left[\varepsilon \log \frac{\varepsilon}{2\beta\sqrt{d}} - \varepsilon \right]_0^{\tau/2} + \frac{\tau}{2} \log(\mathrm{Vol}\,\mathcal{M})
$$

$$
= \frac{d\tau}{2} \left(\log \frac{4\beta\sqrt{d}}{\tau} + 1 \right) + \frac{\tau}{2} \log(\mathrm{Vol}\,\mathcal{M}).
$$

As the covering number is decreasing with increasing ε, the second integral can be bounded as follows:

$$
\int_{\tau/2}^{\rho/2} \log \mathcal{N}(\mathcal{M}, \varepsilon)\, d\varepsilon \le \int_{\tau/2}^{\rho/2} \log \mathcal{N}(\mathcal{M}, \tau/2)\, d\varepsilon
$$

$$
= \left(\frac{\rho}{2} - \frac{\tau}{2} \right) \left[-d \log \frac{\tau}{c\sqrt{d}} + \log(\mathrm{Vol}\,\mathcal{M}) \right]
$$

$$
= d \left(\frac{\rho}{2} - \frac{\tau}{2} \right) \log \frac{c\sqrt{d}}{\tau} + \left(\frac{\rho}{2} - \frac{\tau}{2} \right) \log(\mathrm{Vol}\,\mathcal{M}).
$$

Both together yield

$$
w(\mathcal{M}) \le C \sqrt{\frac{\rho}{2}} \left(\frac{d\tau}{2} \left(\log \frac{c'\sqrt{d}}{\tau} + 1 \right) + d \left(\frac{\rho}{2} - \frac{\tau}{2} \right) \log \frac{c'\sqrt{d}}{\tau} + \frac{\rho}{2} \log(\mathrm{Vol}\,\mathcal{M}) \right)^{1/2}
$$

$$
\le C \sqrt{\frac{\rho}{2}} \left(d\tau \cdot \max\left\{ \log \frac{c'\sqrt{d}}{\tau}, 1 \right\} + d(\rho - \tau) \cdot \max\left\{ \log \frac{c'\sqrt{d}}{\tau}, 1 \right\} + \rho \log(\mathrm{Vol}\,\mathcal{M}) \right)^{1/2}
$$

$$
= C \sqrt{\frac{\rho}{2}} \left(d\rho \cdot \max\left\{ \log \frac{c'\sqrt{d}}{\tau}, 1 \right\} + \rho \log(\mathrm{Vol}\,\mathcal{M}) \right)^{1/2}
$$

$$
\le \frac{C\rho}{\sqrt{2}} \sqrt{ d \cdot \max\left\{ \log \frac{c'\sqrt{d}}{\tau}, 1 \right\} + \log(\mathrm{Vol}\,\mathcal{M}) }.
$$

Appendix C: Proof of Lemmas 4.5 and 4.6

Recall that $\mathscr{M}_j^{\text{rel}} := \{\mathbb{P}_{j,k_j(\mathbf{z})}(\mathbf{z}) : \mathbf{z} \in \mathscr{M}\} \cap B(\mathbf{0}, 2)$. We will begin by establishing some additional technical lemmas.

Lemma C.1 *Set* $C_{\mathscr{M}} := \sup_{\mathbf{z} \in \mathscr{M}} C_{\mathbf{z}}$ *(cf. (iii.b)). Then* $\mathscr{M}_j^{\text{rel}} \subseteq \text{tube}_{C_{\mathscr{M}} 2^{-2j}} \mathscr{M}$.

Proof If $\mathbf{x} \in \mathscr{M}_j^{\text{rel}}$ there exists $\mathbf{z}_{\mathbf{x}} \in \mathscr{M}$ such that $\mathbf{x} = \mathbb{P}_{j,k_j(\mathbf{z}_{\mathbf{x}})}(\mathbf{z}_{\mathbf{x}})$. The Euclidean distance $d(\mathbf{x}, \mathscr{M})$ therefore satisfies

$$d(\mathbf{x}, \mathscr{M}) = \inf_{\mathbf{z} \in \mathscr{M}} \|\mathbf{x} - \mathbf{z}\|_2 \leq \left\|\mathbb{P}_{j,k_j(\mathbf{z}_{\mathbf{x}})}(\mathbf{z}_{\mathbf{x}}) - \mathbf{z}_{\mathbf{x}}\right\|_2 \leq C_{\mathscr{M}} 2^{-2j}$$

by property (iii.b). $\qquad\square$

Given a subset $S \subset \mathbb{R}^D$ we will let $\mathscr{N}(S, \varepsilon)$ denote the cardinality of a minimal ε-cover of S by D-dimensional Euclidean balls of radius $\varepsilon > 0$ each centered in S. Similarly, we will let $\mathscr{P}(S, \varepsilon)$ denote the maximal packing number of S (i.e., the maximum cardinality of a subset of S that contains points all of which are at least Euclidean distance $\varepsilon > 0$ from one another.) The following lemmas bound $\mathscr{N}(\mathscr{M}_j^{\text{rel}}, \varepsilon)$ for various ranges of j and ε.

Lemma C.2 *Set* $C_{\mathscr{M}} := \sup_{\mathbf{z} \in \mathscr{M}} C_{\mathbf{z}}$. *Then* $\mathscr{N}(\mathscr{M}_j^{\text{rel}}, \varepsilon) \leq \mathscr{N}(\mathscr{M}, \varepsilon/2)$ *for all* $\varepsilon \geq 2C_{\mathscr{M}} 2^{-2j}$.

Proof First note that for all $\eta \geq \rho := C_{\mathscr{M}} 2^{-2j}$, Lemma C.1 implies that

$$\mathscr{M}_j^{\text{rel}} \subseteq \text{tube}_\rho \mathscr{M} \subset \bigcup_{\mathbf{p} \in C_\eta} \mathscr{B}(\mathbf{p}, 2\eta),$$

where C_η is an η-cover of $\mathscr{M}$. Thus, for all $\varepsilon \geq 2\eta \geq 2\rho$

$$\mathscr{N}(\mathscr{M}_j^{\text{rel}}, \varepsilon) \leq \mathscr{N}\left(\bigcup_{\mathbf{p} \in C_\eta} \mathscr{B}(\mathbf{p}, 2\eta),\ \varepsilon\right) \leq \mathscr{N}(\mathscr{M}, \eta) = \mathscr{N}(\mathscr{M}, \varepsilon/2). \qquad\square$$

Lemma C.3 $\mathscr{N}(\mathscr{M}_j^{\text{rel}}, \varepsilon) \leq (6/\varepsilon)^d \mathscr{N}(\mathscr{M}, \varepsilon)$ *for all* $\varepsilon \leq C_1 2^{-j}/4$ *as long as* $j > j_0$ *(see properties* (iii.a) *and* (ii.b)*).*

Proof By properties (iii.a) and (ii.b), every center $\mathbf{c}_{j,k}$ has an associated $\mathbf{p}_{j,k} \in \mathscr{M}$ such that both $\mathscr{B}(\mathbf{p}_{j,k}, C_1 2^{-j-2}) \subset \mathscr{B}(\mathbf{c}_{j,k}, C_1 2^{-j-1})$ and $\mathscr{B}(\mathbf{p}_{j,k}, C_1 2^{-j-2}) \cap \mathscr{B}(\mathbf{c}_{j,k'}, C_1 2^{-j-1}) = \emptyset$ for all $k \neq k'$. Let $\tilde{P}_j := \{\mathbf{p}_{j,k} : k \in [K_j]\}$. Consequently, we have that $K_j = |\tilde{P}_j|$ and $\|\mathbf{p}_{j,k} - \mathbf{p}_{j,k'}\|_2 \geq C_1 2^{-j-1}$ for all $k \neq k'$. Since $\tilde{P}_j$ is a $C_1 2^{-j-1}$-packing of $\mathscr{M}$, we can further see that

$$K_j \leq \mathscr{P}(\mathscr{M}, C_1 2^{-j-1}) \leq \mathscr{N}(\mathscr{M}, C_1 2^{-j-2}) \leq \mathscr{N}(\mathscr{M}, \varepsilon)$$

for all $\varepsilon \leq C_1 2^{-j-2}$. Now, $\mathscr{M}_j^{\mathrm{rel}} \subseteq L_j$, where L_j is defined as in the proof of Lemma 4.3 (this proof also discusses its covering numbers). As a result we have that

$$\mathscr{N}(\mathscr{M}_j^{\mathrm{rel}}, \varepsilon) \leq \mathscr{N}(L_j, \varepsilon) \leq K_j \left(\frac{6}{\varepsilon}\right)^d \leq \mathscr{N}(\mathscr{M}, \varepsilon) \cdot \left(\frac{6}{\varepsilon}\right)^d$$

holds for all $\varepsilon \leq C_1 2^{-j-2}$. $\qquad\square$

Furthermore, we will use the two-sided Sudakov inequality as stated in [46].

Lemma C.4 *There exist absolute constants $c, C > 0$ such that the following holds. For $T \in \mathbb{R}^n$, we have that*

$$c \cdot \sup_{\varepsilon \geq 0} \varepsilon \sqrt{\log \mathscr{N}(T, \varepsilon)} \leq w(T) \leq C \log D \cdot \sup_{\varepsilon \geq 0} \varepsilon \sqrt{\log \mathscr{N}(T, \varepsilon)}.$$

C.1: Proof of Lemma 4.5

We aim to bound $w(\mathscr{M}_j^{\mathrm{rel}})$ in terms of $w(\mathscr{M})$. By Lemmas C.1 and C.4, we get that

$$\begin{aligned}
w(\mathscr{M}_j^{\mathrm{rel}}) &\leq C \log D \cdot \sup_{\varepsilon \geq 0} \varepsilon \sqrt{\log \mathscr{N}(\mathscr{M}_j^{\mathrm{rel}}, \varepsilon)} \\
&\leq C \log D \cdot \left(\sup_{0 \leq \varepsilon \leq 2C_{\mathscr{M}} 2^{-2j}} \varepsilon \sqrt{\log \mathscr{N}(\mathrm{tube}_{C_{\mathscr{M}} 2^{-2j}} \mathscr{M}, \varepsilon)} + \sup_{\varepsilon \geq 2C_{\mathscr{M}} 2^{-2j}} \varepsilon \sqrt{\log \mathscr{N}(\mathscr{M}_j^{\mathrm{rel}}, \varepsilon)} \right) \\
&\leq C \log D \cdot \left(\sup_{0 \leq \varepsilon \leq 2C_{\mathscr{M}} 2^{-2j}} \varepsilon \sqrt{\log \mathscr{N}(\mathscr{B}(\mathbf{0}, 1 + C_{\mathscr{M}}), \varepsilon)} + \sup_{\varepsilon \geq 2C_{\mathscr{M}} 2^{-2j}} \varepsilon \sqrt{\log \mathscr{N}(\mathscr{M}, \varepsilon/2)} \right),
\end{aligned}$$

where the last inequality follows from $\mathrm{tube}_{C_{\mathscr{M}} 2^{-2j}} \mathscr{M} \subseteq \mathscr{B}(\mathbf{0}, 1 + C_{\mathscr{M}})$ and Lemma C.2. Appealing to Lemma C.4 once more to bound the second term above we learn that

$$\begin{aligned}
w(\mathscr{M}_j^{\mathrm{rel}}) &\leq C \log D \cdot \left(\sup_{0 \leq \varepsilon \leq 2C_{\mathscr{M}} 2^{-2j}} \varepsilon \sqrt{\log \mathscr{N}(\mathscr{B}(\mathbf{0}, 1 + C_{\mathscr{M}}), \varepsilon))} \right. \\
&\qquad\qquad\qquad \left. + 2 \sup_{\varepsilon \geq 0} \frac{\varepsilon}{2} \sqrt{\log \mathscr{N}(\mathscr{M}, \varepsilon/2)} \right) \\
&\leq C \log D \cdot \left(\sup_{0 \leq \varepsilon \leq 2C_{\mathscr{M}} 2^{-2j}} \varepsilon \sqrt{\log \mathscr{N}(\mathscr{B}(\mathbf{0}, 1 + C_{\mathscr{M}}), \varepsilon))} + 2c w(\mathscr{M}) \right).
\end{aligned}$$

To bound the first term above, we note that using the covering number of $\mathscr{B}(\mathbf{0}, 1 + C_{\mathscr{M}})$ it can be bounded as follows:

$$\begin{aligned}
\mathscr{N}(\mathscr{B}(\mathbf{0}, 1 + C_{\mathscr{M}}), \varepsilon) &= \mathscr{N}\left(\mathscr{B}(\mathbf{0}, 1), \frac{\varepsilon}{1 + C_{\mathscr{M}}}\right) \\
&\leq \left(1 + \frac{2 + 2C_{\mathscr{M}}}{\varepsilon}\right)^D \leq \left(\frac{4C_{\mathscr{M}} + 4}{\varepsilon}\right)^D.
\end{aligned}$$

As $\varepsilon \mapsto \varepsilon\sqrt{\log((4C_\mathscr{M} + 4)/\varepsilon)}$ is non-decreasing for $\varepsilon \in (0, 2C_\mathscr{M}2^{-2j})$, we obtain by assuming that $\log_2 D \leq j$,

$$\sup_{0 \leq \varepsilon \leq 2C_\mathscr{M}\,2^{-2j}} \varepsilon\sqrt{\log \mathscr{N}(\mathscr{B}(\mathbf{0}, 1 + C_\mathscr{M}), \varepsilon))} \leq \sup_{0 \leq \varepsilon \leq 2C_\mathscr{M}\,2^{-2j}} \varepsilon\sqrt{D \log \frac{4C_\mathscr{M} + 4}{\varepsilon}}$$

$$\leq 2C_\mathscr{M}2^{-2j}\sqrt{D \log\left(\frac{4C_\mathscr{M} + 4}{C_\mathscr{M}} \cdot 2^{2j-1}\right)} \leq C_A \cdot \frac{\sqrt{2j-1}}{2^j} \cdot \frac{\sqrt{D}}{2^j} \leq C'$$

where C' is an absolute constant. Appealing to (11) now finishes the proof.

C.2: Proof of Lemma 4.6

Let $2C_\mathscr{M}2^{-2j} \leq \tilde{\varepsilon} \leq C_1 2^{-j}/4$. We aim to bound $w(\mathscr{M}_j^{\mathrm{rel}})$ in terms of covering numbers for $\mathscr{M}$. To do this we will use Dudley's inequality in combination with the knowledge that $\mathscr{M}_j^{\mathrm{rel}} \subset B(\mathbf{0}, 2)$ (by definition). By Dudley's inequality,

$$w(\mathscr{M}_j^{\mathrm{rel}}) \leq C' \int_0^2 \sqrt{\log \mathscr{N}(\mathscr{M}_j^{\mathrm{rel}}, \varepsilon)}\, d\varepsilon$$

$$\leq C'\left(\int_0^{\tilde{\varepsilon}} \sqrt{\log \mathscr{N}(\mathscr{M}_j^{\mathrm{rel}}, \varepsilon)}\, d\varepsilon + \int_{\tilde{\varepsilon}}^2 \sqrt{\log \mathscr{N}(\mathscr{M}_j^{\mathrm{rel}}, \varepsilon)}\, d\varepsilon\right),$$

where C' is an absolute constant. Appealing now to Lemmas C.3 and C.2 for the first and second terms above, respectively, we can see that

$$w(\mathscr{M}_j^{\mathrm{rel}}) \leq C'\left(\int_0^{\tilde{\varepsilon}} \sqrt{\log ((6/\varepsilon)^d \mathscr{N}(\mathscr{M}, \varepsilon))}\, d\varepsilon + \int_{\tilde{\varepsilon}}^2 \sqrt{\log \mathscr{N}(\mathscr{M}, \varepsilon/2)}\, d\varepsilon\right)$$

$$\leq C' \int_0^2 \sqrt{\log ((6/\varepsilon)^d \mathscr{N}(\mathscr{M}, \varepsilon/2))}\, d\varepsilon$$

$$= 2C' \int_0^1 \sqrt{d \log(3/\eta) + \log \mathscr{N}(\mathscr{M}, \eta)}\, d\eta$$

$$\leq 2C'\sqrt{\int_0^1 d \log(3/\eta)\, d\eta + \int_0^1 \log \mathscr{N}(\mathscr{M}, \eta)\, d\eta},$$

where the last bound follows from Jensen's inequality. We can now bound the second term as in the proof of Theorem 3.3 in Appendix B. Doing so we obtain

$$w(\mathscr{M}_j^{\mathrm{rel}}) \leq C''\sqrt{\int_0^1 d \log(3/\eta)\, d\eta + d\left(1 + \log \frac{c'\sqrt{d}}{\tau}\right) + \log(\mathrm{Vol}\,\mathscr{M})}$$

$$\leq C'''\sqrt{d\left(1 + \log \frac{c'\sqrt{d}}{\tau}\right) + \log(\mathrm{Vol}\,\mathscr{M})},$$

where τ is the reach of $\mathscr{M}$ and C''', c' are absolute constants. Appealing to (11) together with Theorem 3.3 now finishes the proof.

Appendix D: Data-Driven GMRA

The axiomatic definition of GMRA proves useful in deducing theoretical results but lacks connection to concrete applications where the structure of $\mathscr{M}$ is not known a priori. Hence, in the following we first describe a probabilistic definition of GMRA which can be well approximated by empirical data (see [3,12,38]) and is connected to the above axioms by applying results from [38]. In fact, we will see that under suitable assumptions the probabilistic GMRA fulfills the axiomatic requirements and its empirical approximation allows one to obtain a version of Theorem 3.1 even when only samples from $\mathscr{M}$ are known.

D.1: Probabilistic GMRA

A probabilistic GMRA of $\mathscr{M}$ with respect to a Borel probability measure Π, as introduced in [38], is a family of (piecewise linear) operators $\{\mathbb{P}_j \colon \mathbb{R}^D \to \mathbb{R}^D\}_{j \geq 0}$ of the form

$$\mathbb{P}_j(\mathbf{x}) = \sum_{k=1}^{K_j} \mathbf{1}_{\mathbf{C_{j,k}}}(\mathbf{x}) \mathbb{P}_{\mathbf{j,k}}(\mathbf{x}).$$

Here, $\mathbf{1_M}$ denotes the indicator function of a set M and, for each refinement level $j \geq 0$, the collection of pairs of measurable subsets and affine projections $\{(C_{j,k}, \mathbb{P}_{j,k})\}_{k=1}^{K_j}$ has the following structure. The subsets $C_{j,k} \subset \mathbb{R}^D$ for $k = 1, \ldots, K_j$ form a partition of $\mathbb{R}^D$, i.e., they are pairwise disjoint and their union is $\mathbb{R}^D$. The affine projectors are defined by

$$\mathbb{P}_{j,k}(\mathbf{x}) = \mathbf{c}'_{j,k} + \mathbb{P}_{V_{j,k}}(\mathbf{x} - \mathbf{c}'_{j,k}),$$

where, for $X \sim \Pi$, $\mathbf{c}'_{j,k} = \mathbb{E}[X | X \in C_{j,k}] =: \mathbb{E}_{j,k}[X] \in \mathbb{R}^D$ and

$$V_{j,k} := \arg\min_{\dim V = d} \mathbb{E}_{j,k}\left[\left\| X - (\mathbf{c}'_{j,k} + \mathrm{Proj}_V(X - \mathbf{c}'_{j,k})) \right\|_2^2 \right],$$

where the minimum is taken over all linear spaces V of dimension d. From now on we will assume uniqueness of these subspaces $V_{j,k}$. To point out parallels to the axiomatic GMRA definition, think of Π being supported on the tube of a d-dimensional manifold. The axiomatic centers $\mathbf{c}_{j,k}$ are then considered to be approximately equal to the conditional means $\mathbf{c}'_{j,k}$ of some cells $C_{j,k}$ partitioning the space, and the corresponding affine projection spaces $P_{j,k}$ are spanned by eigenvectors of the d leading

Table 1 The assumption set on Π

(**A1**) There exists an integer $1 \leq d \leq D$ and a positive constant $\theta_1 = \theta_1(\Pi)$ such that for all $k = 1, \ldots, K_j$,

$\Pi(C_{j,k}) \geq \theta_1 2^{-jd}$.

(**A2**) Define the restricted measure $\Pi_{j,k}$ by $\Pi_{j,k}(S) := \Pi(S \cap C_{j,k})/\Pi(C_{j,k})$ for measurable S. There is a positive constant $\theta_2 = \theta_2(\Pi)$ such that for all $k = 1, \ldots, K_j$, if X is drawn from $\Pi_{j,k}$ then, $\Pi_{j,k}$-almost surely,

$\|X - \mathbf{c}'_{j,k}\|_2 \leq \theta_2 2^{-j}$.

(**A3**) Denote the eigenvalues of the covariance matrix $\Sigma_{j,k}$ by $\lambda_1^{j,k} \geq \ldots \geq \lambda_D^{j,k} \geq 0$. Then there exist $\sigma = \sigma(\Pi) \geq 0$, $\theta_3 = \theta_3(\Pi)$, $\theta_4 = \theta_4(\Pi) > 0$, and some $\alpha > 0$ such that for all $k = 1, \ldots, K_j$,

$\lambda_d^{k,j} \geq \theta_3 \frac{2^{-2j}}{d}$ and $\sum_{l=d+1}^{D} \lambda_l^{j,k} \leq \theta_4(\sigma^2 + 2^{-2(1+\alpha)j}) \leq \frac{\lambda_d^{j,k}}{2}$.

(**A4**) There exists $\theta_5 = \theta_5(\Pi)$ such that

$\|\text{Id} - \mathbb{P}_j\|_{\infty, \Pi} \leq \theta_5(\sigma + 2^{-(1+\alpha)j})$,

where $\|T\|_{\infty, \Pi} = \sup_{x \in \text{supp} \Pi} \|T(x)\|_2$, for $T : \mathbb{R}^D \to \mathbb{R}^D$.

---

eigenvalues of the conditional covariance matrix

$$\Sigma_{j,k} = \mathbb{E}_{j,k}\big[(X - \mathbf{c}'_{j,k})(X - \mathbf{c}'_{j,k})^T\big].$$

Defined in this way, the $\mathbb{P}_j$ correspond to projectors onto the GMRA approximations $\mathcal{M}_j$ introduced above if $\mathbf{c}_{j,k} = \mathbf{c}'_{j,k}$. From [38] we adopt the following assumptions on the entities defined above, and hence, on the distribution Π. From now on we suppose that for all integers $j_{\min} \leq j \leq j_{\max}$, (**A1**)–(**A4**) (see Table 1) hold true.

Remark D.1 Assumption (**A1**) ensures that each partition element contains a reasonable amount of Π-mass. Assumption (**A2**) guarantees that all samples from $\Pi_{j,k}$ will lie close to its expectation/center. As a result, each $\mathbf{c}'_{j,k}$ must be somewhat geometrically central within $C_{j,k}$. Together, (**A1**) and (**A2**) have the combined effect of ensuring that the probability mass of Π is somewhat equally distributed onto the different sets $C_{j,k}$, i.e., the number of points in each set $C_{j,k}$ is approximately the same, at each scale j. The third and fourth assumptions (**A3**) and (**A4**) essentially constrain the geometry of the support of Π to being effectively d-dimensional and somewhat regular (e.g., close to a smooth d-dimensional submanifold of $\mathbb{R}^D$). We refer the reader to [38] for more detailed information regarding these assumptions.

An important class of probability measures Π fulfilling (**A1**)–(**A4**) is presented in [38]. For the sake of completeness we repeat it here and also discuss a method of constructing the partitions $\{C_{jk}\}_{k=1}^{K_j}$ from such probability measures. From here on let $\mathcal{M}$ be a smooth d-dimensional submanifold of $\mathbb{S}^{D-1} \subset \mathbb{R}^D$. Let $\mathcal{U}_K$ denote the uniform distribution on a given set K. We have the following definition.

Definition D.2 ([38, Defn. 3]) Assume that $0 \leq \sigma < \tau$. The distribution Π is said to satisfy the (τ, σ)- *model assumption* if (i) there exists a smooth, compact submanifold $\mathcal{M} \hookrightarrow \mathbb{R}^D$ with reach τ such that supp $\Pi = \text{tube}_\sigma \mathcal{M}$, (ii) the distributions Π and

$\mathscr{U}_{\text{tube}_\sigma\mathscr{M}}$ are absolutely continuous with respect to each other so the Radon–Nikodym derivative $\mathrm{d}\Pi/\mathrm{d}\mathscr{U}_{\text{tube}_\sigma\mathscr{M}}$ exists and satisfies

$$0 < \phi_1 \leq \frac{\mathrm{d}\Pi}{\mathrm{d}\mathscr{U}_{\text{tube}_\sigma\mathscr{M}}} \leq \phi_2 < \infty \qquad \mathscr{U}_{\text{tube}_\sigma(\mathscr{M})}\text{-almost surely.}$$

The constants ϕ_1 and ϕ_2 are implicitly assumed to only depend on a slowly growing function of D, compare [38, Rem. 4].

Let us now discuss the construction of suitable partitions $\{C_{jk}\}$ by making use of cover trees. A *cover tree* T on a finite set of samples $S \subset \mathscr{M}$ is a hierarchy of levels with the starting level containing the root point and the last level containing every point in S. To every level a set of nodes is assigned which is associated with a subset of points in S. To be precise, given a set S of n distinct points in some metric space $(\mathbb{X}, d_{\mathbb{X}})$, a cover tree T on S is a sequence of subsets $T_i \subset S$, $i = 0, 1, \ldots$, that satisfies the following (see [8]).

(a) **Nesting:** $T_i \subseteq T_{i+1}$, i.e., once a point appears in T_i it is in every T_j for $j \geq i$.
(b) **Covering:** For every $\mathbf{x} \in T_{i+1}$ there exists exactly one $\mathbf{y} \in T_i$ such that $d_{\mathbb{X}}(\mathbf{x}, \mathbf{y}) \leq 2^{-i}$. Here $\mathbf{y}$ is called the *parent* of $\mathbf{x}$.
(c) **Separation:** For all distinct points $\mathbf{x}, \mathbf{y} \in T_i$, $d_{\mathbb{X}}(\mathbf{x}, \mathbf{y}) > 2^{-i}$.

The set T_i denotes the set of points in S associated with nodes at level i. Note that there exists $N \in \mathbb{N}$ such that $T_i = S$ for all $i \geq N$. Herein we will presume that S is large enough to contain an ϵ-cover of $\mathscr{M}$ for $\epsilon > 0$ sufficiently small.

Moreover, the axioms characterizing cover trees are strongly connected to the dyadic structure of GMRA. For a given cover tree (for construction see [8]) on a set $\mathscr{X}_n = \{X_1, \ldots, X_n\}$ of i.i.d. samples from the distribution Π with respect to the Euclidean distance let $\mathbf{a}_{j,k}$ for $k = 1, \ldots, K_j$ be the elements of the jth level of the cover tree, i.e., $T_j = \{\mathbf{a}_{j,k}\}_{k=1}^{K_j}$, and define

$$\kappa_j(x) = \underset{1 \leq k \leq K_j}{\arg\min} \|\mathbf{x} - \mathbf{a}_{j,k}\|_2.$$

With this a partition of $\mathbb{R}^D$ into *Voronoi regions*

$$C_{j,k} = \{\mathbf{x} \in \mathbb{R}^D : \kappa_j(\mathbf{x}) = k\} \tag{20}$$

can be defined. Maggioni et al. showed in [38, Thm. 7] that by this construction all assumptions (**A1**)–(**A4**) can be fulfilled. The question arises if the properties of the axiomatic definition of GMRA in Definition 2.1 are equally met. As only parts of the axioms are relevant for our analysis, we refrain from giving rigorous justification for all properties.

1. GMRA property (i) holds by construction if the matrices $\Phi_{j,k}$ are defined so that $\Phi_{j,k}^T \Phi_{j,k} = \mathbb{P}_{V_{j,k}}$ along with any reasonable choice of centers $\mathbf{c}_{j,k}$.

2. The dyadic structure axioms (ii.a)–(ii.c) also hold as a trivial consequence of the cover tree properties (a)–(c) above if the axiomatic centers $\mathbf{c}_{j,k}$ are chosen to be the elements of the cover tree set T_j (i.e., the $\mathbf{a}_{j,k}$ elements). By the (ρ, σ)-model assumption, samples drawn from Π will have a quite uniform distribution all over supp Π. Hence, the probabilistic centers $\mathbf{c}'_{j,k}$ of each $C_{j,k}$-set will also tend to be close to the axiomatic centers $\mathbf{c}_{j,k} = \mathbf{a}_{j,k}$ proposed here for small σ (see, e.g., assumption (**A2**) above).

3. One can deduce GMRA property (iii.a) from the fact that our chosen centers $\mathbf{a}_{j,k}$ belong to $\mathcal{M}$ if supp $\Pi = \mathcal{M}$ (or to a small tube around $\mathcal{M}$ if σ is small).

4. The first part of (iii.b) is implied by (**A4**) with the uniform constant θ_5 for all $\mathbf{x} \in \mathcal{M}$ if $\mathbf{a}_{j,k}$ is sufficiently close to $\mathbf{c}'_{j,k}$. To show the second part of (iii.b) note that

$$
\begin{aligned}
\|\mathbf{x} - \mathbb{P}_{j,k'}(\mathbf{x})\|_2 &\leq \|\mathbf{x} - \mathbf{c}_{j,k'}\|_2 + \|\mathbf{c}_{j,k'} - \mathbb{P}_{j,k'}(\mathbf{x})\|_2 \\
&= \|\mathbf{x} - \mathbf{c}_{j,k'}\|_2 + \big\|\mathbb{P}_{V_{j,k'}}(\mathbf{x} - \mathbf{c}_{j,k'})\big\|_2 \\
&\leq 2\|\mathbf{x} - \mathbf{c}_{j,k'}\|_2 \leq 32 \max\big\{\big\|\mathbf{x} - \mathbf{c}_{j,k_j(\mathbf{x})}\big\|_2, C_1 2^{-j-1}\big\} \\
&\leq 32 \max\{C_\epsilon 2^{-j}, C_1 2^{-j-1}\} \leq C \cdot 2^{-j}
\end{aligned}
$$

where in the second last step, we used our cover tree properties (recall that $\mathbf{c}_{j,k} = \mathbf{a}_{j,k}$). Again, the constants $C, C_\epsilon > 0$ do not depend on the chosen $x \in \mathcal{M}$ as long as S is well chosen (e.g., contains a sufficiently fine cover of $\mathcal{M}$).

Considering the GMRA axioms above we can now see that only the first part of (iii.b) may not hold in a satisfactory manner if we choose to set $\Phi_{j,k}^T \Phi_{j,k} = \mathbb{P}_{V_{j,k}}$ and $\mathbf{c}_{j,k} = \mathbf{a}_{j,k}$. And, even when it does not hold with C_z being independent of j, it will at least hold with a worse j-dependence due to assumption (**A2**).

D.2: Empirical GMRA

The axiomatic properties only hold above, of course, if the GMRA is constructed with knowledge of the true $\mathbb{P}_{V_{j,k}}$ subspaces. In reality, however, this will not be the case and we are rather given some training data consisting of n samples from near/on $\mathcal{M}$, $\mathscr{X}_n = \{X_1, \ldots, X_n\}$, which we assume to be i.i.d. with distribution Π. These samples are used to approximate the real GMRA subspaces based on Π so that the operators $\mathbb{P}_j$ can be replaced by their estimators

$$
\widehat{\mathbb{P}}_j(\mathbf{x}) = \sum_{k=1}^{K_j} \mathbf{1}_{\{\mathbf{x} \in \mathbf{C}_{j,k}\}} \widehat{\mathbb{P}}_{j,k}(\mathbf{x}),
$$

where $\{C_{j,k}\}_{k=1}^{K_j}$ is a suitable partition of $\mathbb{R}^d$ obtained from the data,

$$
\widehat{\mathbb{P}}_{j,k}(\mathbf{x}) = \widehat{\mathbf{c}}_{j,k} + \mathbb{P}_{\widehat{V}_{j,k}}(\mathbf{x} - \widehat{\mathbf{c}}_{j,k}), \qquad \widehat{\mathbf{c}}_{j,k} = \frac{1}{|\mathscr{X}_{k,j}|} \sum_{\mathbf{x} \in \mathscr{X}_{j,k}} \mathbf{x},
$$

$$\widehat{V}_{j,k} = \underset{\dim V = d}{\arg\min} \frac{1}{|\mathscr{X}_{j,k}|} \sum_{\mathbf{x} \in \mathscr{X}_{j,k}} \|\mathbf{x} - \widehat{\mathbf{c}}_{j,k} - \mathbb{P}_V(\mathbf{x} - \widehat{\mathbf{c}}_{j,k})\|_2^2,$$

and $\mathscr{X}_{j,k} = C_{j,k} \cap \mathscr{X}_n$. In other words, working with the above model we have one perfect GMRA that cannot be computed (unless Π is known) but fulfills all important axiomatic properties, and an estimated GMRA that is at hand but that is only an approximation to the perfect one. Thankfully, the main results of [38] stated in Appendix E give error bounds on the difference between perfect and estimated GMRA with $\mathbf{c}_{j,k} = \widehat{\mathbf{c}}_{j,k} \approx \mathbf{c}'_{j,k} \approx \mathbf{a}_{j,k}$ that only depend on the number of samples from Π one can acquire. Following their notational convention we will denote the empirical GMRA approximation at level j, i.e., the set $\widehat{\mathbb{P}}_j$ projects onto, by $\widehat{\mathscr{M}_j} = \{\widehat{\mathbb{P}}_j(\mathbf{z}) : \mathbf{z} \in \mathscr{B}(\mathbf{0}, 2)\} \cap \mathscr{B}(\mathbf{0}, 2)$ and the affine subspaces by $\widehat{P}_{j,k} = \{\widehat{\mathbb{P}}_{j,k}(\mathbf{z}) : \mathbf{z} \in \mathbb{R}^D\}$. We again restrict the approximation to $\mathscr{B}(\mathbf{0}, 2)$. The single affine spaces will be non-empty as all $\widehat{\mathbf{c}}_{j,k}$ lie by definition close to $\mathscr{B}(\mathbf{0}, 1)$ if supp Π is close to $\mathscr{M}$, which we assume.

In the empirical setting OMS has to be slightly modified to conform to our empirical GMRA notation. Hence, (6) and (7) become

$$\widehat{\mathbf{c}}_{j,k'} \in \underset{\widehat{\mathbf{c}}_{j,k} \in \widehat{\mathscr{C}_j}}{\arg\min} \; d_{\mathrm{H}}(\mathrm{sign}\, A\widehat{\mathbf{c}}_{j,k}, \mathbf{y}), \tag{21}$$

$$\mathbf{x}^* = \underset{\mathbf{z} \in \mathbb{R}^D}{\arg\min} \sum_{l=1}^m (-y_l)\langle \mathbf{a}_l, \mathbf{z}\rangle \text{ subject to } \mathbf{z} \in \mathrm{conv}\, \mathbb{P}_{\mathbb{S}}(\widehat{P}_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2)). \tag{22}$$

OMS can be adapted in a similar way by changing (6) and (7). To stay consistent with the axiomatic notation, we denote the sets containing the centers $\mathbf{c}'_{j,k}$ and $\widehat{\mathbf{c}}_{j,k}$ by $\mathscr{C}'_j$ and $\widehat{\mathscr{C}_j}$, respectively. As shown in Appendix E, the main result also holds in this setting. There is only an additional influence of sample size on the probability.

Appendix E: Proof of Theorem 3.1 with Empirical GMRA

Recall the definitions of probabilistic GMRA, empirical GMRA, and the modifications of (6) resp. (7) to become (21) resp. (22). We start with the central result of [38].

Theorem E.1 *Suppose that assumptions* **(A1)–(A3)** *are satisfied (see Table 1). Let* $X, X_1, \ldots, X_n$ *be an i.i.d. sample from* Π *and* $\bar{d} = 4d^2\theta_2^4/\theta_3^2$. *Then for any* $j_{\min} \leq j \leq j_{\max}$ *and any* $t \geq 1$ *such that* $t + \log\max\{\bar{d}, 8\} \leq \theta_1 n 2^{-jd}/2$,

$$\mathbb{E}\big[\|X - \widehat{\mathbb{P}}_j(X)\|_2^2\big] \leq 2\theta_4(\sigma^2 + 2^{-2j(1+\alpha)}) + c_1 2^{-2j} \frac{d^2(t + \log\max\{\bar{d}, 8\})}{n 2^{-jd}},$$

and if in addition **(A4)** *is satisfied,*

$$\|\mathrm{Id} - \widehat{\mathbb{P}}_j\|_{\infty, \Pi} \leq \theta_5(\sigma + 2^{-(1+\alpha)j}) + \sqrt{\frac{c_1}{2} 2^{-2j} \frac{d^2(t + \log\max\{\bar{d}, 8\})}{n 2^{-jd}}}$$

with probability $\geq 1 - 2^{jd+1}(e^{-t} + e^{-\theta_1 n 2^{-jd}}/16)/\theta_1$, *where* $c_1 = 2(12\theta_2^3\sqrt{2}/(\theta_3\sqrt{\theta_1}) + 4\theta_2\sqrt{2}/(d\sqrt{\theta_1}))^2$.

Theorem E.1 states that under assumptions (**A1**)–(**A4**) the empirical GMRA approximates $\mathcal{M}$ as well as the perfect probabilistic one as long as the number of samples n is sufficiently large. For the proof of our main theorem we only need the following two bounds which can be deduced from (20) and (21) in [38] by setting $t = 2^{jd}$. As both appear in the proof of Theorem E.1, we state them as a corollary. The interested reader may note that $n_{j,k}$ appearing in the original statements can be lower bounded by $\theta_1 n 2^{-jd}$.

Corollary E.2 *Under the assumptions of Theorem E.1 the following holds for any* $C_1 > 0$ *as long as* j, α *are sufficiently large and* σ *is sufficiently small:*

$$\Pr\left[\max_{k\in K_j}\left\|\mathbb{P}_{V_{j,k}} - \mathbb{P}_{\widehat{V}_{j,k}}\right\| \geq \frac{C_1}{12}2^{-j-2}\right] \leq \frac{2}{\theta_2}2^{jd}e^{-2^{jd}\min\{1, 32\theta_2^2 d^2/C_1^2\}},$$

$$\Pr\left[\max_{k\in K_j}\left\|\mathbf{c}'_{j,k} - \widehat{\mathbf{c}}_{j,k}\right\|_2 \geq \frac{C_1}{12}2^{-j-2}\right] \leq \frac{2}{\theta_2}2^{jd}e^{-2^{jd}\min\{1, 32\theta_2^2 d^2/C_1^2\}},$$

if

$$n \geq n_{\min} = (2^{jd} + \log\max\{\bar{d}, 8\})\cdot$$
$$\cdot\min\left\{144\cdot 2^{(d+1)j+3}\theta_2^2 d/(C_1\theta_1\theta_3), 96\cdot 2^{dj+1}\theta_2/(C_1\theta_1)\right\}^2.$$

Remark E.3 By Corollary E.2 with probability of at least $1 - \mathcal{O}(2^{jd}e^{-2^{jd}})$ the empirical centers $\widehat{\mathbf{c}}_{j,k}$ of one level j have a worst-case distance to the perfect centers $\mathbf{c}'_{j,k}$ of at most $\mathcal{O}(2^{-j-2})$ if $n \gtrsim \mathcal{O}(2^{3jd})$. As a result, the empirical centers $\widehat{\mathbf{c}}_{j,k}$ will also be at most $\mathcal{O}(2^{-j-2})$ distance from their associated cover tree centers $\mathbf{a}_{j,k}$ if $n \gtrsim \mathcal{O}(2^{3jd})$, by assumption (**A2**). The same holds true for the projectors $\mathbb{P}_{V_{j,k}}$ and $\mathbb{P}_{\widehat{V}_{j,k}}$ in operator norm.

The proof of Theorem 3.1 in this setting follows the same steps as in the axiomatic one. First, we give an empirical version of Lemma 4.13. Then we link $\mathbf{x}$ and $\mathbf{x}^*$ as described in Sect. 4.2 while controlling the difference between empirical and axiomatic but unknown GMRA by Corollary E.2. The following extension of Lemma 4.3 will be regularly used.

Corollary E.4 (bound of Gaussian width) *The Gaussian width of* $\mathcal{M} \cup \mathcal{M}_j \cup \widehat{\mathcal{M}}_j$ *can be bounded from above by*

$$\max\{w(\mathcal{M}), w(\mathcal{M}_j), w(\widehat{\mathcal{M}}_j)\} \leq w(\mathcal{M} \cup \mathcal{M}_j \cup \widehat{\mathcal{M}}_j)$$
$$\leq 2w(\mathcal{M}) + 2w(\mathcal{M}_j) + 2w(\widehat{\mathcal{M}}_j) + 5 \leq 2w(\mathcal{M}) + C\sqrt{dj}$$

where $\widehat{\mathcal{M}}_j$ *is defined as at the end of Appendix D.*

Proof The proof follows directly the lines of the proof of Lemma 4.3. The additional term $w(\widehat{\mathcal{M}}_j)$ can be bounded in the same way as $w(\mathcal{M}_j)$.   □

Remark E.5 By the structure of the proof one can easily obtain several subversions of the inequalities, e.g., $w(\mathcal{M} \cup \widehat{\mathcal{M}}_j) \leq 2w(\mathcal{M}) + 2w(\widehat{\mathcal{M}}_j) + 5$. We will use them while only referring to Corollary E.4. Moreover, similar generalizations as in Lemma 4.3 apply (cf. Remark 4.4)

Note that we are now setting our empirical GMRA centers $\mathbf{c}_{j,k}$ to be the associated mean estimates $\widehat{\mathbf{c}}_{j,k}$ as a means of approximating the axiomatic GMRA structure we would have if we had instead chosen our centers to be the true expectations $\mathbf{c}'_{j,k}$ (recall Appendix D). We also implicitly assume below that there exists a constant $C_1 > 0$ for which the associated axiomatic GMRA properties in Sect. 2 hold when the centers $\mathbf{c}_{j,k}$ are chosen as these true expectations $\mathbf{c}'_{j,k}$ and the $\Phi_{j,k}^T \Phi_{j,k}$ as $\mathbb{P}_{V_{j,k}}$.

Lemma E.6 *Fix j sufficiently large. Under the assumptions of Theorem E.1 and $n \geq n_{\min}$, if $m \geq \bar{C} C_1^{-6} 2^{6(j+1)} w(\mathcal{M} \cup \mathbb{P}_{\mathbb{S}}(\widehat{\mathscr{C}}_j))^2$ the index k' of the center $\widehat{\mathbf{c}}_{j,k'}$ chosen in step I of the algorithm fulfills*

$$\left\| \mathbf{x} - \mathbf{c}'_{j,k'} \right\|_2 \leq 16 \max \left\{ \left\| \mathbf{x} - \mathbf{c}'_{j,k_j(\mathbf{x})} \right\|_2, C_1 2^{-j-1} \right\}$$

for all $\mathbf{x} \in \mathcal{M} \subset \mathbb{S}^{D-1}$ with probability at least $1 - \mathcal{O}\big(2^{jd} e^{-2^{jd}} + e^{\delta^2 m}\big)$.

Proof The proof will be similar to the one of Lemma 4.13. By definition we have

$$d_{\mathrm{H}}(\mathrm{sign}\, A\widehat{\mathbf{c}}_{j,k'}, \mathbf{y}) \leq d_{\mathrm{H}}(\mathrm{sign}\, A\widehat{\mathbf{c}}_{j,k_j(\mathbf{x})}, \mathbf{y}).$$

As, for all $\mathbf{z}, \mathbf{z}' \in \mathbb{S}^{D-1}$, $d_{\mathrm{H}}(\mathrm{sign}\, A\mathbf{z}, \mathrm{sign}\, A\mathbf{z}') = m \cdot d_A(\mathbf{z}, \mathbf{z}')$, this is equivalent to

$$d_A(\mathbb{P}_{\mathbb{S}}(\widehat{\mathbf{c}}_{j,k'}), \mathbf{x}) \leq d_A(\mathbb{P}_{\mathbb{S}}(\widehat{\mathbf{c}}_{j,k_j(\mathbf{x})}), \mathbf{x}).$$

Theorem 4.9 transfers the bound to normalized geodesic distance, namely

$$d_{\mathrm{G}}(\mathbb{P}_{\mathbb{S}}(\widehat{\mathbf{c}}_{j,k'}), \mathbf{x}) \leq d_{\mathrm{G}}(\mathbb{P}_{\mathbb{S}}(\widehat{\mathbf{c}}_{j,k_j(\mathbf{x})}), \mathbf{x}) + 2\delta$$

with probability at least $1 - 2e^{-c\delta^2 m}$ where $\delta = C_1 2^{-j-1}$. Observe $d_{\mathrm{G}}(\mathbf{z}, \mathbf{z}') \leq \|\mathbf{z} - \mathbf{z}'\|_2 \leq \pi\, d_{\mathrm{G}}(\mathbf{z}, \mathbf{z}')$ for all $\mathbf{z}, \mathbf{z}' \in \mathbb{S}^{D-1}$ (see Lemma 4.7) which leads to

$$\|\mathbb{P}_{\mathbb{S}}(\widehat{\mathbf{c}}_{j,k'}) - \mathbf{x}\|_2 \leq \pi\, d_{\mathrm{G}}(\mathbb{P}_{\mathbb{S}}(\widehat{\mathbf{c}}_{j,k_j(\mathbf{x})}), \mathbf{x}) + 2\pi\delta \leq \pi\, \|\mathbb{P}_{\mathbb{S}}(\widehat{\mathbf{c}}_{j,k_j(\mathbf{x})}) - \mathbf{x}\|_2 + 2\pi\delta.$$

We will now use the fact that by Corollary E.2,

$$\|\widehat{\mathbf{c}}_{j,k} - \mathbf{c}'_{j,k}\|_2 \leq \frac{C_1}{12} 2^{-j-2}$$

for all $k \in K_j$ with probability $\geq 1 - \mathcal{O}\big(2^{jd} e^{-2^{jd}}\big)$. From this we first deduce by GMRA property (iii.a) that $\|\widehat{\mathbf{c}}_{j,k} - \mathbb{P}_{\mathbb{S}}(\widehat{\mathbf{c}}_{j,k})\|_2 \leq \|\widehat{\mathbf{c}}_{j,k} - \mathbb{P}_{\mathbb{S}}(\mathbf{c}'_{j,k})\|_2 \leq \|\widehat{\mathbf{c}}_{j,k} - \mathbf{c}'_{j,k}\|_2$

$+ \|\mathbf{c}'_{j,k} - \mathbb{P}_{\mathbb{S}}(\mathbf{c}'_{j,k})\|_2 < (C_1 + C_1/2)2^{-j-2}$ for all $\widehat{\mathbf{c}}_{j,k} \in \widehat{\mathscr{C}}_j$. Combining above estimates and using triangle inequality we obtain

$$
\begin{aligned}
\|\mathbf{c}_{j,k'} - \mathbf{x}\|_2 &\leq \|\mathbb{P}_{\mathbb{S}}(\widehat{\mathbf{c}}_{j,k'}) - \mathbf{x}\|_2 + \|\widehat{\mathbf{c}}_{j,k'} - \mathbb{P}_{\mathbb{S}}(\widehat{\mathbf{c}}_{j,k'})\|_2 + \|\widehat{\mathbf{c}}_{j,k'} - \mathbf{c}'_{j,k'}\|_2 \\
&< \pi \|\mathbb{P}_{\mathbb{S}}(\widehat{\mathbf{c}}_{j,k_j(x)}) - \mathbf{x}\|_2 + 2\pi\delta + 2C_1 2^{-j-2} \\
&\leq \pi \left(\|\widehat{\mathbf{c}}_{j,k_j(\mathbf{x})} - \mathbb{P}_{\mathbb{S}}(\widehat{\mathbf{c}}_{j,k_j(\mathbf{x})})\|_2 + \|\widehat{\mathbf{c}}_{j,k_j(\mathbf{x})} - \mathbf{c}'_{j,k_j(\mathbf{x})}\|_2 + \|\mathbf{c}'_{j,k_j(\mathbf{x})} - \mathbf{x}\|_2 \right) \\
&\quad + 2\pi\delta + C_1 2^{-j-1} < \pi \|\mathbf{c}_{j,k_j(\mathbf{x})} - \mathbf{x}\|_2 + 2\pi\delta + (1+\pi)C_1 2^{-j-1} \\
&\leq (4\pi + 1) \max\{\|c_{j,k_j(\mathbf{x})} - \mathbf{x}\|_2, C_1 2^{-j-1}\} \\
&\leq 16 \max\{\|c_{j,k_j(\mathbf{x})} - \mathbf{x}\|_2, C_1 2^{-j-1}\}.
\end{aligned}
$$

A union bound over both probabilities yields the result. □

Having Lemma E.6 at hand we can now show a detailed version of Theorem 3.1 in this case. For convenience please first read the proof of Theorem 4.14. As above choosing $\varepsilon = 2^{-j}\sqrt{j}$ yields Theorem 3.1 for OMS-simple with a slightly modified probability of success and slightly different dependencies on C_1 and $\tilde{C}_{\mathbf{x}}$ in (9).

Theorem E.7 *Let $\mathscr{M} \subset \mathbb{S}^{D-1}$ be given by its empirical GMRA for some levels $j_0 \leq j \leq J$ from samples $X_1, \ldots, X_n$ for $n \geq n_{\min}$ (defined in Corollary E.2), such that $0 < C_1 < 2^{j_0+1}$ where C_1 is the constant from GMRA properties* (ii.b) *and* (iii.a) *for a GMRA structure constructed with centers $\mathbf{c}'_{j,k}$ and with the $\Phi^T_{j,k}\Phi_{j,k}$ as $\mathbb{P}_{V_{j,k}}$. Fix j and assume that $\mathrm{dist}(\mathbf{0}, \widehat{\mathscr{M}_j}) \geq 1/2$. Further, let*

$$
m \geq 64 \max\{C', \bar{C}\} \cdot C_1^{-6} 2^{6(j+1)}(w(\mathscr{M}) + C\sqrt{dj})^2,
$$

where C' is the constant from Theorem 4.11, $\bar{C}$ from Theorem 4.9, and C from Lemma 4.3. Then, with probability at least $1 - \mathscr{O}(2^{jd}e^{-2^{jd}} + e^{\delta^2 m})$ the following holds for all $\mathbf{x} \in \mathscr{M}$ with one-bit measurements $y = \mathrm{sign}\, A\mathbf{x}$ and GMRA constants $\tilde{C}_{\mathbf{x}}$ from property (iii.b) *satisfying $\tilde{C}_{\mathbf{x}} < 2^{j-1}$: The approximations $\mathbf{x}^*$ obtained by OMS fulfill*

$$
\begin{aligned}
\|\mathbf{x} - \mathbf{x}^*\|_2 \leq \Bigg(&2^{1-j/2}\left(\tilde{C}_{\mathbf{x}} + \frac{C_1}{8}\right) + \sqrt{C_1}\sqrt[4]{\log \frac{4e}{C_1}} \\
&+ \sqrt{22\tilde{C}_{\mathbf{x}} + \frac{55}{4}C_1}\sqrt[4]{\log \frac{2e}{2\tilde{C}_{\mathbf{x}} + 5C_1/4}} \Bigg) 2^{-j/2}\sqrt[4]{j}.
\end{aligned}
$$

Proof The proof consists of the same three steps as the one of Theorem 4.14. First, we apply Lemma E.6 in **(I)**. By the GMRA axioms this supplies an estimate for $\|\mathbf{x} - \mathbb{P}_{j,k'}(\mathbf{x})\|_2$ with high probability (recall that $\mathbb{P}_{j,k'}(\mathbf{x})$ will be $\mathbb{P}_{V_{j,k'}}(\mathbf{x} - \mathbf{c}'_{j,k'}) + \mathbf{c}'_{j,k'}$ in this case). In **(II)** we use **(I)** to deduce a bound on $\|\mathbf{x} - \widehat{\mathbb{P}}_{j,k'}(\mathbf{x})\|_2$, and then use

Theorem 4.11 to bound the distance between $\widehat{\mathbb{P}}_{j,k'}(\mathbf{x})/\|\widehat{\mathbb{P}}_{j,k'}(\mathbf{x})\|_2$ and the minimum point $\mathbf{x}^*$ of

$$\mathbf{x}^* = \arg\min_{\mathbf{z}} \sum_{l=1}^{m} (-y_l)\langle \mathbf{a}_l, \mathbf{z} \rangle \quad \text{subject to } \mathbf{z} \in K := \text{conv}\, \mathbb{P}_{\mathbb{S}}(\widehat{P}_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2)), \quad (23)$$

with high probability. Taking the union bound over all events, part **(III)** then concludes with an estimate of the distance $\|\mathbf{x} - \mathbf{x}^*\|_2$ by combining **(I)** and **(II)**.

(I) Set $\delta = C_1 2^{-j-1}$ and recall that $C_1 2^{-j-2} < 1/2$ by assumption which implies by GMRA property (iii.a) that all centers in $\mathscr{C}'_j$ are closer to $\mathbb{S}^{D-1}$ than $1/2$, i.e., $1/2 \leq \|\mathbf{c}'_{j,k}\|_2 \leq 3/2$. Moreover, Corollary E.2 holds with probability at least $1 - \mathscr{O}(2^{jd}e^{-2^{jd}})$ and implies $\|\widehat{\mathbf{c}}_{j,k} - \mathbf{c}'_{j,k}\|_2 \leq (C_1/12)2^{-j-2} \leq 1/4$. Hence, by triangle inequality, $1/4 \leq \|\widehat{\mathbf{c}}_{j,k}\|_2 \leq 7/4$. From this and (10) we deduce

$$w(\mathbb{P}_{\mathbb{S}}(\widehat{\mathscr{C}}_j)) \leq \gamma(\mathbb{P}_{\mathbb{S}}(\widehat{\mathscr{C}}_j)) \leq 4\gamma(\widehat{\mathscr{C}}_j) \leq 8w(\widehat{\mathscr{C}}_j) + 4\sqrt{\frac{2}{\pi}}\,\text{dist}(\mathbf{0}, \widehat{\mathscr{C}}_j) \leq 8w(\widehat{\mathscr{C}}_j) + 8. \quad (24)$$

As $\widehat{\mathscr{C}}_j \subset \widehat{\mathscr{M}}_j$, we know by Corollary E.4 and (24) that

$$\begin{aligned}
m &\geq 16\bar{C}\delta^{-6}(2w(\mathscr{M}) + 2C\sqrt{dj})^2 \geq 16\bar{C}\delta^{-6}(2w(\mathscr{M}) + 4w(\widehat{\mathscr{C}}_j) + 10)^2 \\
&\geq 4\bar{C}\delta^{-6}(4w(\mathscr{M}) + 8w(\widehat{\mathscr{C}}_j) + 20)^2 \geq 4\bar{C}\delta^{-6}(4w(\mathscr{M}) + w(\mathbb{P}_{\mathbb{S}}(\widehat{\mathscr{C}}_j)) + 12)^2 \\
&\geq \bar{C}\delta^{-6}(8w(\mathscr{M}) + 2w(\mathbb{P}_{\mathbb{S}}(\widehat{\mathscr{C}}_j)) + 24)^2 \geq \bar{C}\delta^{-6}(w(\mathscr{M} \cup \mathbb{P}_{\mathbb{S}}(\widehat{\mathscr{C}}_j)) + 19)^2 \\
&\geq \bar{C}\delta^{-6} \max\left\{ w(\mathscr{M} \cup \mathbb{P}_{\mathbb{S}}(\widehat{\mathscr{C}}_j))^2, \frac{2}{\pi} \right\}. \quad (25)
\end{aligned}$$

Hence, Lemma 4.13 implies

$$\|\mathbf{x} - \mathbf{c}'_{j,k'}\|_2 \leq 16 \max\left\{ \left\|\mathbf{x} - \mathbf{c}'_{j,k_j(\mathbf{x})}\right\|_2, C_1 2^{-j-1} \right\}$$

with probability at least $1 - \mathscr{O}(2^{jd}e^{-2^{jd}} + e^{\delta^2 m})$. By GMRA property (iii.b) we get, for some constant $\tilde{C}_{\mathbf{x}}$,

$$\|\mathbf{x} - \mathbb{P}_{j,k'}(\mathbf{x})\|_2 \leq \tilde{C}_{\mathbf{x}} 2^{-j}. \quad (26)$$

(II) Define $\widehat{\alpha} = \|\widehat{\mathbb{P}}_{j,k'}(\mathbf{x})\|_2$. Note that $\|\mathbf{x} - \mathbf{c}'_{j,k'}\|_2 \leq 4$ as $\mathbf{x} \in \mathbb{S}^{D-1}$ and all $\mathbf{c}'_{j,k}$ are close to the sphere by assumption. Hence,

$$\begin{aligned}
&\|\mathbb{P}_{j,k'}(\mathbf{x}) - \widehat{\mathbb{P}}_{j,k'}(\mathbf{x})\|_2 \\
&\quad \leq \left\|\mathbf{c}'_{j,k'} + \mathbb{P}_{V_{j,k'}}(\mathbf{x} - \mathbf{c}'_{j,k'}) - \widehat{\mathbf{c}}_{j,k'} - \mathbb{P}_{\widehat{V}_{j,k'}}(\mathbf{x} - \widehat{\mathbf{c}}_{j,k'})\right\|_2 \\
&\quad \leq \|\mathbf{c}'_{j,k'} - \widehat{\mathbf{c}}_{j,k'}\|_2 + \left\|\mathbb{P}_{V_{j,k'}} - \mathbb{P}_{\widehat{V}_{j,k'}}\right\| \cdot \|\mathbf{x} - \mathbf{c}'_{j,k'}\|_2 + \|\mathbf{c}'_{j,k'} - \widehat{\mathbf{c}}_{j,k'}\|_2
\end{aligned}$$

$$\leq \frac{2C_1 2^{-j-2}}{12} + \frac{C_1 2^{-j-2}}{12} \|\mathbf{x} - \mathbf{c}'_{j,k'}\|_2 \leq \frac{C_1 2^{-j-2}}{2},$$

by application of Corollary E.2. This implies $1/4 \leq \widehat{\alpha} \leq 7/4$ as $\mathbf{x} \in \mathbb{S}^{D-1}$ and

$$\|\mathbf{x} - \widehat{\mathbb{P}}_{j,k'}(\mathbf{x})\|_2 \leq \|\mathbf{x} - \mathbb{P}_{j,k'}(\mathbf{x})\|_2 + \|\mathbb{P}_{j,k'}(\mathbf{x}) - \widehat{\mathbb{P}}_{j,k'}(\mathbf{x})\|_2$$
$$\leq \tilde{C}_{\mathbf{x}} 2^{-j} + \frac{C_1 2^{-j-2}}{2} \leq \frac{3}{4} \tag{27}$$

by (26) and the assumption that $\max\{\tilde{C}_{\mathbf{x}}, C_1/4\} \cdot 2^{-j} \leq 1/2$. As before we create the setting for Theorem 4.11.

Let us define $\tilde{\mathbf{x}} := \widehat{\mathbb{P}}_{j,k'}(\mathbf{x})/\widehat{\alpha} \in \mathbb{S}^{D-1}$, $\tilde{\mathbf{y}} := \operatorname{sign} A\tilde{\mathbf{x}} = \operatorname{sign} A\widehat{\mathbb{P}}_{j,k'}(\mathbf{x})$, $K = \operatorname{conv} \mathbb{P}_{\mathbb{S}}(\widehat{P}_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2))$, and $\tau := 2^{-j}(2\tilde{C}_{\mathbf{x}} + 5C_1/4)$. If applied to this, Theorem 4.11 would give the desired bound on $\|\tilde{\mathbf{x}} - \mathbf{x}^*\|_2$. We first have to check $d_{\mathrm{H}}(\tilde{\mathbf{y}}, \mathbf{y}) \leq \tau m$. Recall that $1/\widehat{\alpha} \leq 4$ and as $\widehat{\alpha} > 0$ one has $\widehat{\alpha} w(K) = w(\widehat{\alpha} K)$. By applying Corollary E.4 again we have that

$$m \geq 64\delta^{-6} \max\{C', \bar{C}\} \cdot (w(\mathscr{M}) + C\sqrt{dj})^2$$
$$\geq 4\bar{C}\delta^{-6}(2w(\mathscr{M}) + 2w(\widehat{\mathscr{M}_j}) + 5)^2$$
$$\geq \bar{C}\delta^{-6}(8w(\mathscr{M}) + 8w(\widehat{\mathscr{M}_j}) + 20)^2$$
$$\geq \bar{C}\delta^{-6} w\left(\mathscr{M} \cup \frac{\widehat{\mathscr{M}_j}}{\widehat{\alpha}}\right)^2 \geq \bar{C}\delta^{-6} w\left(\left(\mathscr{M} \cup \frac{\widehat{\mathscr{M}_j}}{\widehat{\alpha}}\right) \cap \mathscr{B}(\mathbf{0}, 1)\right)^2.$$

Use $\|\tilde{\mathbf{x}} - \widehat{\mathbb{P}}_{j,k'}(\mathbf{x})\|_2 = |1 - \widehat{\alpha}| \leq \|\mathbf{x} - \widehat{\mathbb{P}}_{j,k'}(\mathbf{x})\|_2$, Theorem 4.9, and Lemma 4.7, to obtain

$$\frac{d_{\mathrm{H}}(\tilde{\mathbf{y}}, \mathbf{y})}{m} = d_A(\tilde{\mathbf{x}}, \mathbf{x}) \leq d_G(\tilde{\mathbf{x}}, \mathbf{x}) + 2\delta \leq \|\tilde{\mathbf{x}} - \mathbf{x}\|_2 + 2\delta$$
$$\leq \|\tilde{\mathbf{x}} - \widehat{\mathbb{P}}_{j,k'}(\mathbf{x})\|_2 + \|\widehat{\mathbb{P}}_{j,k'}(\mathbf{x}) - \mathbf{x}\|_2 + 2\delta \leq 2\|\widehat{\mathbb{P}}_{j,k'}(\mathbf{x}) - \mathbf{x}\|_2 + 2\delta$$
$$\leq 2\tilde{C}_{\mathbf{x}} 2^{-j} + C_1 2^{-j-2} + 2\delta \leq 2^{-j}\left(2\tilde{C}_{\mathbf{x}} + \frac{5C_1}{4}\right) = \tau$$

with probability at least $1 - 2e^{-c\delta^2 m}$. Assuming the above events hold true we can apply Theorem 4.11, as by Corollary E.4, in analogy to (25) and (18),

$$m \geq 4C'\delta^{-6}(2w(\mathscr{M}) + 4w(\mathscr{M}_j) + 4w(\widehat{\mathscr{M}_j}) + 10)^2$$
$$\geq C'\delta^{-6} w\left(\mathbb{P}_{\mathbb{S}}(\widehat{P}_{j,k'} \cap \mathscr{B}(\mathbf{0}, 2))\right) \geq C'\delta^{-6} w(K)^2$$

and obtain with probability at least $1 - 8e^{-c\delta^2 m}$,

$$\|\tilde{\mathbf{x}} - \mathbf{x}^*\|_2^2 \leq \delta\sqrt{\log \frac{e}{\delta}} + 11\tau\sqrt{\log \frac{e}{\tau}}. \tag{28}$$

(III) We conclude as in Theorem 4.14. Recall that $\|\tilde{\mathbf{x}} - \widehat{\mathbb{P}}_{j,k'}(\mathbf{x})\|_2 = |1 - \alpha| \leq \|\mathbf{x} - \mathbb{P}_{j,k'}(\mathbf{x})\|_2 \leq 2^{-j}(\tilde{C}_{\mathbf{x}} + C_1/8)$. By union bound we obtain with probability at least $1 - \mathcal{O}\big(2^{jd}e^{-2^{jd}} + e^{\delta^2 m}\big)$,

$$
\begin{aligned}
\|\mathbf{x} - \mathbf{x}^*\|_2 &\leq \|\mathbf{x} - \widehat{\mathbb{P}}_{j,k'}(\mathbf{x})\|_2 + \|\widehat{\mathbb{P}}_{j,k'}(\mathbf{x}) - \tilde{\mathbf{x}}\|_2 + \|\tilde{\mathbf{x}} - \mathbf{x}^*\|_2 \\
&\leq 2\|\mathbf{x} - \widehat{\mathbb{P}}_{j,k'}(\mathbf{x})\|_2 + \sqrt{\delta\sqrt{\log\frac{e}{\delta}} + 11\tau\sqrt{\log\frac{e}{\tau}}} \\
&\leq 2^{1-j}\left(\tilde{C}_{\mathbf{x}} + \frac{C_1}{8}\right) + 2^{-(j+1)/2}\sqrt{C_1}\sqrt[4]{\log\frac{e}{C_1 2^{-j-1}}} \\
&\quad + 2^{-j/2}\sqrt{22\tilde{C}_{\mathbf{x}} + 55C_1/4}\sqrt[4]{\log\frac{2^j e}{2\tilde{C}_{\mathbf{x}} + 5C_1/4}} \\
&\leq 2^{-j/2}\sqrt[4]{j}\left\{2^{1-j/2}\left(\tilde{C}_{\mathbf{x}} + \frac{C_1}{8}\right) + \sqrt{C_1}\sqrt[4]{\log\frac{4e}{C_1}}\right. \\
&\quad \left. + \sqrt{22\tilde{C}_{\mathbf{x}} + \frac{55}{4}C_1}\sqrt[4]{\log\frac{2e}{2\tilde{C}_{\mathbf{x}} + 5C_1/4}}\right\}.
\end{aligned}
$$

As explained in the proof of Theorem 4.14, the last step was simplified for notational reasons. $\qquad\square$

References

1. Ahmed, A., Recht, B., Romberg, J.: Blind deconvolution using convex programming. IEEE Trans. Inform. Theory **60**(3), 1711–1732 (2014)
2. Ai, A., Lapanowski, A., Plan, Y., Vershynin, R.: One-bit compressed sensing with non-Gaussian measurements. Linear Algebra Appl. **441**, 222–239 (2014)
3. Allard, W.K., Chen, G., Maggioni, M.: Multi-scale geometric methods for data sets II: geometric multi-resolution analysis. Appl. Comput. Harmon. Anal. **32**(3), 435–462 (2012)
4. Bandyopadhyay, S., Giannella, C., Maulik, U., Kargupta, H., Liu, K., Datta, S.: Clustering distributed data streams in peer-to-peer environments. Inform. Sci. **176**(14), 1952–1985 (2006)
5. Baraniuk, R.G., Wakin, M.B.: Random projections of smooth manifolds. Found. Comput. Math. **9**(1), 51–77 (2009)
6. Benedetto, J.J., Powell, A.M., Yılmaz, Ö.: Sigma-Delta ($\Sigma\Delta$) quantization and finite frames. IEEE Trans. Inform. Theory **52**(5), 1990–2005 (2006)
7. Benedetto, J.J., Powell, A.M., Yılmaz, Ö.: Second-order sigma-delta ($\Sigma\Delta$) quantization of finite frame expansions. Appl. Comput. Harmon. Anal. **20**(1), 126–148 (2006)
8. Beygelzimer, A., Kakade, S., Langford, J.: Cover trees for nearest neighbor. In: 23rd International Conference on Machine Learning (Pittsburgh 2006), pp. 97–104. ACM, New York (2006)
9. Boufounos, P.T., Jacques, L., Krahmer, F., Saab, R.: Quantization and compressive sensing. In: Compressed Sensing and its Applications. Appl. Numer. Harmon. Anal., pp. 193–237. Springer, Cham (2015)
10. Candès, E.J., Romberg, J.K., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Commun. Pure Appl. Math. **59**(8), 1207–1223 (2006)
11. Candès, E.J., Strohmer, T., Voroninski, V.: PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming. Commun. Pure Appl. Math. **66**(8), 1241–1274 (2013)

12. Chen, G., Iwen, M., Chin, S., Maggioni, M.: A fast multiscale framework for data in high-dimensions: measure estimation, anomaly detection, and compressive measurements. In: Visual Communications and Image Processing (San Diego 2012). IEEE, New York (2013)
13. Dirksen, S., Iwen, M., Krause-Solberg, S., Maly, J.: Robust one-bit compressed sensing with manifold data. In: 13th International Conference on Sampling Theory and Applications (Bordeaux 2019). IEEE, New York (2019)
14. Dirksen, S., Jung, H.Ch., Rauhut, H.: One-bit compressed sensing with partial Gaussian circulant matrices. Inf. Inference **9**(3), 601–626 (2020)
15. Donoho, D.L.: Compressed sensing. IEEE Trans. Inform. Theory **52**(4), 1289–1306 (2006)
16. Dudley, R.M.: V.N. Sudakov's work on expected suprema of Gaussian processes. In: High Dimensional Probability VII (Cargèse 2014). Progr. Probab., vol. 71, pp. 37–43. Springer, Cham (2016)
17. Eftekhari, A., Wakin, M.B.: New analysis of manifold embeddings and signal recovery from compressive measurements. Appl. Comput. Harmon. Anal. **39**(1), 67–109 (2015)
18. Eldar, Y.C., Mishali, M.: Robust recovery of signals from a structured union of subspaces. IEEE Trans. Inform. Theory **55**(11), 5302–5316 (2009)
19. Federer, H.: Curvature measures. Trans. Am. Math. Soc. **93**(3), 418–491 (1959)
20. Feng, J., Krahmer, F.: An RIP-based approach to $\Sigma\Delta$ quantization for compressed sensing. IEEE Signal Process. Lett. **21**(11), 1351–1355 (2014)
21. Feng, J.M., Krahmer, F., Saab, R.: Quantized compressed sensing for partial random circulant matrices. Appl. Comput. Harmon. Anal. **47**(3), 1014–1032 (2019)
22. Gray, R.: Oversampled sigma-delta modulation. IEEE Trans. Commun. **35**(5), 481–489 (1987)
23. Güntürk, C.S., Lammers, M., Powell, A.M., Saab, R., Yılmaz, Ö.: Sobolev duals for random frames and $\Sigma\Delta$ quantization of compressed sensing measurements. Found. Comput. Math. **13**(1), 1–36 (2013)
24. Haghighatshoar, S., Caire, G.: Low-complexity massive MIMO subspace estimation and tracking from low-dimensional projections. IEEE Trans. Signal Process. **66**(7), 1832–1844 (2018)
25. Huynh, T., Saab, R.: Fast binary embeddings and quantized compressed sensing with structured matrices. Commun. Pure Appl. Math. **73**(1), 110–149 (2020)
26. Indyk, P., Motwani, R.: Approximate nearest neighbors: towards removing the curse of dimensionality. In: 30th Annual ACM Symposium on Theory of Computing (Dallas 1998), pp. 604–613. ACM, New York (1999)
27. Iwen, M.A., Krahmer, F.: Fast subspace approximation via greedy least-squares. Constr. Approx. **42**(2), 281–301 (2015)
28. Iwen, M.A., Lybrand, E., Nelson, A.A., Saab, R.: New algorithms and improved guarantees for one-bit compressed sensing on manifolds. In: 13th International Conference on Sampling Theory and Applications (Bordeaux 2019). IEEE, New York (2019)
29. Iwen, M.A., Maggioni, M.: Approximation of points on low-dimensional manifolds via random linear projections. Inf. Inference **2**(1), 1–31 (2013)
30. Iwen, M.A., Ong, B.W.: A distributed and incremental SVD algorithm for agglomerative data analysis on large networks. SIAM J. Matrix Anal. Appl. **37**(4), 1699–1718 (2016)
31. Jacques, L., Laska, J.N., Boufounos, P.T., Baraniuk, R.G.: Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. IEEE Trans. Inform. Theory **59**(4), 2082–2102 (2013)
32. Jung, H.Ch., Maly, J., Palzer, L., Stollenwerk, A.: Quantized compressed sensing by rectified linear units. Proc. Appl. Math. Mech. (2021). https://doi.org/10.1002/pamm.202000015
33. Krahmer, F., Saab, R., Yılmaz, Ö.: Sigma-Delta quantization of sub-Gaussian frame expansions and its application to compressed sensing. Inf. Inference **3**(1), 40–58 (2014)
34. Krause-Solberg, S., Maly, J.: A tractable approach for one-bit Compressed Sensing on manifolds. In: 12th International Conference on Sampling Theory and Applications (Tallinn 2017), pp. 667–671. IEEE, New York (2017)
35. Latorre, F., Eftekhari, A., Cevher, V.: Fast and provable ADMM for learning with generative priors. In: 33rd Conference on Neural Information Processing Systems (Vancouver 2019), pp. 12027–12039. Curran Associates, Red Hook (2019)
36. LeCun, Y., Cortes, C., Burges, Ch.J.C.: THE MNIST DATABASE of handwritten digits. http://yann.lecun.com/exdb/mnist/
37. Liao, W., Maggioni, M.: Adaptive geometric multiscale approximations for intrinsically low-dimensional data. J. Mach. Learn. Res. **20**, # 98 (2019)
38. Maggioni, M., Minsker, S., Strawn, N.: Multiscale dictionary learning: non-asymptotic bounds and robustness. J. Mach. Learn. Res. **17**, # 2 (2016)

39. Mondal, B., Dutta, S., Heath Jr., R.W.: Quantization on the Grassmann manifold. IEEE Trans. Signal Process. **55**(8), 4208–4216 (2007)
40. Norsworthy, S.R., Schreier, R., Temes, G.C.: Delta-Sigma-Converters. Design and Simulation. IEEE, New York (1996)
41. Plan, Y., Vershynin, R.: Robust 1-bit compressed sensing and sparse logistic regression: a convex programming approach. IEEE Trans. Inform. Theory **59**(1), 482–494 (2013)
42. Plan, Y., Vershynin, R.: One-bit compressed sensing by linear programming. Commun. Pure Appl. Math. **66**(8), 1275–1297 (2013)
43. Plan, Y., Vershynin, R.: Dimension reduction by random hyperplane tessellations. Discrete Comput. Geom. **51**(2), 438–461 (2014)
44. Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev. **52**(3), 471–501 (2010)
45. Saab, R., Wang, R., Yılmaz, Ö.: Quantization of compressive samples with stable and robust recovery. Appl. Comput. Harmon. Anal. **44**(1), 123–143 (2018)
46. Vershynin, R.: High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge Series in Statistical and Probabilistic Mathematics, vol. 47. Cambridge University Press, Cambridge (2018)
47. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms (2017). arXiv:1708.07747

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.