New Algorithms and Improved Guarantees for One-Bit Compressed Sensing on Manifolds

Mark A. Iwen†, Eric Lybrand‡, Aaron A. Nelson‡*, Rayan Saab‡

Abstract—We study the problem of approximately recovering signals on a manifold from one-bit linear measurements drawn from either a Gaussian ensemble, partial circulant ensemble, or bounded orthonormal ensemble and quantized using $\Sigma\Delta$ or distributed noise shaping schemes. We assume we are given a Geometric Multi-Resolution Analysis, which approximates the manifold, and we propose a convex optimization algorithm for signal recovery. We prove an upper bound on the recovery error which outperforms prior works that use memoryless scalar quantization, requires a simpler analysis, and extends the class of measurements beyond Gaussians. Finally, we illustrate our results with numerical experiments.

Index Terms—Compressed sensing, quantization, one-bit, manifold, rate-distortion

I. INTRODUCTION

Compressed sensing [3], [5] demonstrates that structured high dimensional signals such as sparse vectors or low-rank matrices can be recovered from few random linear measurements. Recovery is typically formulated as a convex optimization problem whose minimizer cannot be expressed analytically and must be solved for using numerical algorithms running on digital devices. Thus, it is necessary to consider the effect of quantization in the design of the recovery algorithms. Indeed, sparse vector recovery and low-rank matrix recovery have been studied in the presence of various quantization schemes [7], [8], [11], [12], [13]. We look to extend these results to account for those structured signals that lie on a compact, low-dimensional submanifold of $\mathbb{R}^N$ for which we have a Geometric Multi-Resolution Analysis (GMRA) [1]. Our work is motivated by the results of Iwen et al. in [9] where they assume memoryless scalar quantized Gaussian measurements, and we provide better error bounds that hold for a wider class of measurement ensembles.

As in [9], a key component of our technique is the GMRA which approximates the manifold at various levels of refinement. At each level the GMRA is a collection of approximate tangent spaces about certain known "centers", and the quality of the approximation improves with every level. Unlike in [9], the quantization schemes that we use are $\Sigma\Delta$ or distributed noise shaping methods (see, e.g., [7], [8]) and the compressed sensing measurements that our results apply to include those drawn from Gaussian ensembles, partial circulant ensembles (PCE) or bounded orthonormal ensembles (BOE) (see [8] for precise definitions). Our proposed reconstruction method is summarized in Algorithm 1. This simple algorithm first finds a GMRA center that quantizes to a bit sequence close to the quantized measurements, where "closeness" is determined using a pseudometric that respects the quantization; it then optimizes over all points in the associated approximate tangent space to enforce, as much as possible, the consistency of the quantization. Using the results of [8] we prove that the quantization error associated with our proposed reconstruction algorithm decays polynomially or exponentially as a function of the number of measurements, depending on the quantization scheme. This greatly improves on the sub-linear error decay associated with scalar quantization in [9].

II. BACKGROUND

In [9], Iwen et al. study the case where measurements $y = Ax$ of a signal $x$ on a manifold $K \subset S^{N-1}$ are quantized via memoryless scalar quantization (MSQ). For a discrete set $\mathcal{A}$ and $\mathcal{Q}(x) := \arg\min_{z \in \mathcal{A}} |x - z|$, the measurements are

$$q_j = \mathcal{Q}(\langle a_j, x \rangle), \quad j = 1, \ldots, m.$$
For example, one could take $A = \{\pm 1\}$ and $\varnothing(\cdot) = \text{sign}(\cdot)$. [9] proposes an algorithm for recovering $x$ from such measurements and shows that the associated error decays like $O(m^{-1/2})$. Such slow error decay, associated with MSQ, has also been seen in the context of sparse vector recovery in the compressed sensing literature. Indeed, it is known in that setting that the error under any reconstruction scheme using MSQ measurements cannot decay faster than $O(m^{-1})$ [6] (see also [2]). So, to achieve better error rates one must use more sophisticated quantization schemes. For example, in the sparse vector setting noise shaping techniques such as $\Sigma\Delta$ and distributed noise shaping leverage redundancy of the measurements to ensure error decay like $O(m^{-r})$ or $O(\beta^{-cmn})$ for some parameters $r \in \mathbb{N}$, $\beta > 1$ that depend on the quantization scheme, e.g., [4], [14]. As we will also use these schemes, we now briefly describe them.

Each of the quantization methods mentioned above employs a state variable $u \in \mathbb{R}^m$ and quantizes measurements in a recursive fashion: $q_j = \varnothing\{f(y_j, \ldots, y_1, u_{j-1}, \ldots, u_1)\}$, where $f$ is some function designed for the quantization scheme. The state variable is then updated via the state relation $Ax - q = Hu$, where $H : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is a lower-triangular Toeplitz matrix. Important for the analysis (and for practical reasons) is that $H, f$ are chosen so that whenever $\|x\|_\infty$ is bounded, we have $\|u\|_\infty < C$; the upper bound $C$ is often referred to as the stability constant of the quantization scheme. For a more detailed explanation of these noise shaping techniques, the interested reader may refer for example to [8]. For the sake of expositional simplicity, we will only consider the $\Sigma\Delta$ setting, but our arguments work for distributed noise shaping under very minor adjustments.

III. PROBLEM FORMULATION AND NOTATION

For an integer $\ell$, $|\ell| = \{1, \ldots, \ell\}$. We use $\geq$ and $\leq$ for inequalities that hold up to a constant; subscripts indicate the constant depends on a specified parameter. Let $K \subset B_2^N$ be a $d$-dimensional submanifold of the unit $\ell_2$-ball in $\mathbb{R}^N$. We assume that we have a GMRA of $K$, which we make precise below. First, for a set $T \subset \mathbb{R}^N$ and $\rho > 0$, define

$$\text{tube}_\rho(T) := \{x \in \mathbb{R}^N : \inf_{y \in T} \|x - y\|_2 \leq \rho\}.$$ 

**Definition 1** ([10]). Let $j \in \mathbb{N}$ and $K_0, \ldots, K_j \in \mathbb{N}$. A GMRA of $K$ is a collection $\{\mathcal{E}_j, \mathcal{P}_j\}_{j \in |J|}$ of centers $\mathcal{E}_j = \{c_{j,k} : k \in [K_j]\}$ and affine projections

$$\mathcal{P}_j = \{P_{j,k} : \mathbb{R}^N \rightarrow \mathbb{R}^N : k \in [K_j]\}$$

with the following properties:

1) **Affine Projections.** Every $P_{j,k}$ is an orthogonal projection onto some $d$-dimensional affine space which contains the center $c_{j,k}$.

2) **Dyadic Structure.** The number of centers at each level is bounded by $|\mathcal{E}_j| = K_j \leq C_\Sigma 2^dj$ for an absolute constant $C_\Sigma \geq 1$. Moreover, there exist $C_1 > 0$, $C_2 \in (0, 1)$ such that

   a) $K_j \leq K_{j+1}$ for all $j \in |J| - 1$,

   b) $\|c_{j,k} - c_{j,k}\|_2 > C_1 2^{-j}$ for all $j \in |J|$, $k_1 \neq k_2 \in [K_j]$.

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

   $$\|c_{j,k} - c_{j-1,p_j(k)}\|_2 \leq C_2 \min_{k' \in [K_{j-1}]} \|c_{j,k} - c_{j-1,k'}\|_2.$$ 

3) **Multiscale Approximation.** The projectors in $\mathcal{P}_j$ approximate $K$ in the following sense:

   a) There exists $j_0 \in |J| - 1$ such that $c_{j,k} \in \text{tube}_{G_{j-2^{j-2}}}(K)$ for all $j \geq j_0$ and $k \in [K_j]$.

   b) For each $j \in |J|$ and $z \in \mathbb{R}^N$, let

   $$c_{j,k}(z) \in \arg \min_{c_{j,k} \in \mathcal{E}_j} \|z - c_{j,k}\|_2.$$ 

   Then for each $z \in K$ there exist $C_z, \tilde{C}_z > 0$ so that

   $$\|z - P_{j,k}(z)\|_2 \leq C_z 2^{-j}$$

   for all $j \in |J|$ and

   $$\|z - P_{j,k}(z)\|_2 \leq \tilde{C}_z 2^{-j}$$

   whenever $j \in |J|$ and $k' \in [K_j]$ satisfy

   $$\|z - c_{j,k'}\|_2 \leq 16 \max \{\|z - c_{j,k}(z)\|_2, G_{j-2^{j-1}}\}.$$ 

Let $\{\mathcal{E}_j, \mathcal{P}_j\}_{j \in J}$ be a GMRA of a smooth compact manifold $K \subset (1 - \mu)B_2^N$ for some $\mu \in (0, 1)$, and define the scale-$j$ GMRA approximation $\mathcal{K}_j := \{P_{j,k}(z) : \|z\|_2 \leq 2^j\} \subset B_2^N$. We suppose that $j_0$ is large enough so that $\sup_{x \in K} \tilde{C}_x 2^{-j_0} \leq \mu$ to ensure $P_{j,k}(x) : x \in K, (j, k')$ as in 3.b of Definition 1$\subset B_2^N$, and further assume that $\text{tube}_{G_{j-2^{j-2}}}(K) \subset B_2^N$ which ensures $\mathcal{E}_j \subset \mathcal{K}_j$ for $j \geq j_0$. The number of measurements required for our theoretical guarantees to hold will depend on two notions of complexity of $K$ and the GMRA. For $g \sim N(0, I_N)$, define

$$w(S) := \mathbb{E}\sup_{x \in \mathcal{S}} \langle g, x \rangle, \quad \text{rad}(S) := \sup_{x \in \mathcal{S}} \|x\|_2,$$

and for $j \geq j_0$, define

$$C_K := \max \{C_1, \sup_{x \in K} \tilde{C}_x\}, \quad S := K \cup \mathcal{K}_j.$$ 

Now, let $\varnothing$ be a stable $r^{th}$ order $\Sigma\Delta$ quantizer with stability constant $C(r)$ and associated alphabet $A$. Let $x \in K, A \subset \mathbb{R}^{m \times N}$ be a standard Gaussian matrix (or a matrix drawn from a PCE/BOE), $D \in \mathbb{R}^{N \times N}$ a
diagonal matrix with random signs (independent of \( A \)) along the diagonal, \( \Phi := AD \), and

\[
q := \mathcal{Q}(\Phi x).
\]

Our goal, given \( q \) and \( \Phi \), is to approximate \( x \) and show that the associated error bounds decay fast as a function of \( m \).

A useful fact is that the binary embedding provided by \( \Sigma \Delta \) quantization approximately preserves Euclidean distance via a related pseudo-metric on the quantized vectors, defined as follows. For \( p \leq m \), define \( \lambda := m/p = r\hat{\lambda} - r + 1 \) and \( v \in \mathbb{R}^\lambda \) be the (row) vector whose \( j \)th entry \( v_j \) is the \( j \)th coefficient of the polynomial of \((1 + z + \ldots + z^\lambda)^r\). Set \( \gamma = \|v\|_1 / \|v\|_2 \) and define \( \tilde{V} : \mathbb{R}^m \to \mathbb{R}^p \) by

\[
\tilde{v} = \frac{1}{\sqrt{\|v\|_2}} I_p \otimes v,
\]

where \( \otimes \) denotes the Kronecker product. Then the pseudo-metric is given by \( d_\tilde{V}(x, y) := \|\tilde{V}(x - y)\|_2 \).

IV. MAIN RESULTS

We now present our recovery algorithm and its associated error guarantees.

**Algorithm 1** Reconstruction Algorithm

1. **Step 1:** Find \( c_{j,k} \in \arg\min_{z} \|\tilde{V}(\mathcal{Q}(c_{j,k} x) - q)\|_2 \).
2. **Step 2:** If \( \|\tilde{V}(\mathcal{Q}(c_{j,k} x) - q)\|_2 = 0 \), set \( x^\dagger = c_{j,k} \); else

\[
x^\dagger = \arg\min_{z \in \mathbb{R}^N} \|\tilde{V}(\Phi z - q)\|_2 \quad \text{s.t.} \quad z = P_{j,k}(z), \quad \|z\|_2 \leq 1.
\]

**Theorem 2.** Suppose we have a GMRA of \( K \subset (1 - \mu)B_2^N \) at level \( j \geq j_0 \). Let \( \mathcal{Q} \) be a stable \( r \)th order \( \Sigma \Delta \) quantizer with \( r = O(j) \), associated alphabet \( \mathcal{A} = \{\pm(1 - \alpha)\sqrt{m} \} \) where \( \alpha^{-1} \gtrsim \text{rad}(S - S)2^{2(j+1)} \), and

\[
m^{r - 1/2} \gtrsim \gamma^2 (1 + \nu)^2 \log^4(N) 2^{24j} \text{rad}(S - S) \max\{1, w^2(S - S) \text{rad}^{-2}(S - S)\}.
\]

Then with probability exceeding \( 1 - e^{-\nu} \), for all \( x \in K \), \( x^\dagger \) from Algorithm 1 satisfies

\[
\|x^\dagger - x\|_2 \lesssim \tilde{C}_\lambda 2^{-j}.
\]

**Proof.** By the triangle inequality we have

\[
\|x^\dagger - x\|_2 \leq \|x^\dagger - P_{j,k} x\|_2 + \|P_{j,k} x - x\|_2.
\]

Both terms are bounded by \( \tilde{C}_\lambda 2^{-j} \), the first by Lemma 4 and the second by Lemma 3 and Equation (1) from Definition 1.

**Remark 1.** As Lemma 4.3 of [9] shows, \( w(S - S) \leq w(K) + \sqrt{\lambda j} \). This is a suitable bound for coarse GMRA scales, i.e. \( j \lesssim \log(N) \). However, for \( j \gtrsim \log(N) \) one can slightly modify the definition of \( S \) and use the bound \( w(S - S) \leq (w(K) + 1) \log(N) \) as proven in Lemma 4.5 of [9], albeit this requires some modifications to the proof of Theorem 2. Please see Remark 4.15 of [9] for more details, as we shall leave the latter case for future work.

**Lemma 3.** Under the assumptions of Theorem 2, with probability exceeding \( 1 - e^{-\nu} \), for all \( x \in K \), the center \( c_{j,k} \) chosen in Step 1 of Algorithm 1 satisfies

\[
\|x - c_{j,k}\|_2 \leq 16 \max\{\|x - c_{j,k}(x)\|_2, C_1 \cdot 2^{j-1}\}.
\]

**Proof.** Theorem 5.2 and Remarks 3, 5 of [8] state that if \( m^{r - 1/2} \geq \gamma^2 (1 + \nu)^2 \log^4(N) \max\{1, w^2(S - S) \text{rad}^{-2}(S - S)\} \cdot \frac{a^2}{\lambda} \) and we choose \( r = \frac{1}{\sqrt{\log(N)}} \), where \( \lambda = m^{r - 1/2} / p \) and \( C > 0 \), then with probability exceeding \( 1 - e^{-\nu} \)

\[
\|d_\mathcal{Q}(\mathcal{Q}(c_{j,k}), q) - \|c_{j,k} - x\|_2\|_2 \lesssim \max \{\sqrt{\alpha}, \alpha\} \text{rad}(S - S) + e^{-c\sqrt{\lambda}}.
\]

for all \( c_{j,k} \in \mathcal{C}_j \). Conditioning on this, we have

\[
d_\mathcal{Q}(\mathcal{Q}(c_{j,k}), q) \lesssim \|c_{j,k} - x\|_2 \quad \text{for all} \quad c_{j,k} \in \mathcal{C}_j.
\]

**Lemma 4.** Under the assumptions of Theorem 2, with probability exceeding \( 1 - e^{-\nu} \), for all \( x \in K \), \( x^\dagger \) from Algorithm 1 satisfies

\[
\|x^\dagger - P_{j,k} x\|_2 \lesssim \tilde{C}_\lambda 2^{-j}.
\]

**Proof.** By optimality of \( x^\dagger \) and feasibility of \( P_{j,k} x \), the triangle inequality gives us

\[
0 \leq \|\tilde{V}(\Phi x^\dagger - x)\|_2 \leq \|\tilde{V}(\Phi x - q)\|_2 \leq 2 \|\tilde{V}(\Phi P_{j,k} x - q)\|_2 + 2d_\mathcal{Q}(q, q^\star).
\]

Define \( q^\star := \mathcal{Q}(\Phi P_{j,k} x) \). Then

\[
\|\tilde{V}(\Phi x^\dagger - P_{j,k} x)\|_2 \leq 2 \|\tilde{V}(\Phi P_{j,k} x - q^\star + q^\star - q)\|_2 + 2d_\mathcal{Q}(q, q^\star).
\]

Lemma 4.5 of [8] states that \( \|\tilde{V}(\Phi P_{j,k} x - q^\star)\|_2 \lesssim e^{-c\sqrt{\lambda}} 2^{-j} \), while Theorem 5.2 of [8] and (1) (via Lemma 3) imply \( d_\mathcal{Q}(q, q^\star) \lesssim \tilde{C}_\lambda 2^{-j} \) with probability exceeding \( 1 - e^{-\nu} \). Therefore, we have

\[
\|\tilde{V}(\Phi x^\dagger - P_{j,k} x)\|_2 \lesssim e^{-c\sqrt{\lambda}} 2^{-j},
\]
By the definition of \( S \), we have \( x^i, P_{j,k} x \in S \). Equation (5.8) in the proof of Theorem 5.2 of [8] states that with probability exceeding \( 1 - e^{-v} \|	ilde{V}\Phi(x^i - P_{j,k} x)\|_2 \|x^i - P_{j,k} x\|_2 2^{-j} \). With (4), this yields \( \|x^i - P_{j,k} x\|_2 \leq C_x 2^{-j} \).

\[ \square \]

Remark 2. If one does not choose \( r \) and \( \alpha \) as in the proofs above, but works with \( r \) large enough and \( \alpha \) small enough, a version of Theorem 2 with \( \|x^i - x\|_2 \leq C_x 2^{-j} + \max(\sqrt{\alpha}, \alpha) \text{rad}(S - S) + (m/p)^{-r + 1/2} \) holds for \( m \), \( p \) as in the proof of Lemma 3.

Remark 3. When \( \Sigma\Delta \) quantization is replaced by distributed noise shaping, Theorem 2 holds with \( j2^{4j} \) in the lower bound (2) replaced by \( j2^{4j} \), and Remark 2 holds with \( \beta^{-m/p} \) replacing \( (m/p)^{-r + 1/2} \).

V. Numerical Simulations

To simulate Algorithm 1, we take \( K = \mathbb{S}^2 \) embedded in \( \mathbb{R}^{20} \) and construct a GMRA up to level \( j_{\text{max}} = 15 \) using 20,000 data points sampled uniformly from \( K \). We randomly select a test set of 100 points \( x \in K \) for use throughout all experiments. In each experiment (i.e., point in Figure 1), compressed sensing measurements \( y = \Phi x = m^{-1/2} AD_c x \) are taken for each test point, with \( A \sim \mathcal{N}(0, I_{m \times N}) \) and \( D_c \) a diagonal \( N \times N \) matrix of random \( \pm 1 \)s. We recover \( x^i \) from the \( r \)th order \( \Sigma\Delta \) measurements \( \mathcal{Q}(y) \) via Algorithm 1 where, for practical reasons, the alphabet from Theorem 2 is modified to be \( \mathcal{A} = \{ \pm 1 \} \). We vary \( \lambda = m/p \) for fixed \( r \), \( p \), and refinement scale \( j \). As in Remark 2, the reconstruction error decays as a function of \( \lambda \) until reaching a floor due to the refinement level of the GMRA.

VI. Acknowledgements

We thank F. Krahmer, S. Krause-Solberg, and J. Maly for sharing their GMRA code, which they adapted from that provided by M. Maggioni. M. A. Iwen was supported in part by NSF CCF-1615489; R. Saab was supported in part by NSF DMS-1517204.

References