Improved Recovery Guarantees for One-Bit Compressed Sensing on Manifolds

Eric Lybrand

May 22, 2019

UC San Diego

Coauthors

Mark Iwen, MSU

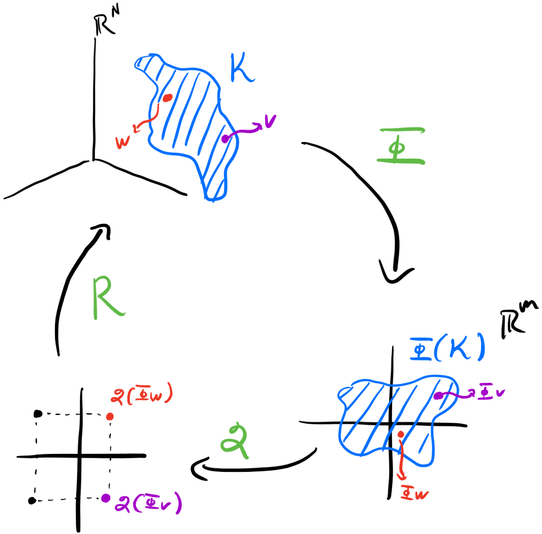
Aaron Nelson, UCSD

Rayan Saab, UCSD

Introduction

- Why compressed sensing?
  - Data acquisition technique that simultaneously reduces dimension.
  - Many useful tools for fast linear near-isometric embeddings, avoids curse of dimensionality, ... (more later)
- Why quantization?
  - Compressed sensing algorithms require using digital computers ... (more later)
  - Reduces memory overhead associated with high dimensional data.
- Why manifolds?
  - Useful model for data in signal processing, machine learning.

The Mental Picture



Wish List

Embedding

- Must be fast, e.g. convolutions, DFT, DCT.
- Approximately preserves structure of data, e.g. pairwise distances.

Wish List

Embedding

- Must be fast, e.g. convolutions, DFT, DCT.
- Approximately preserves structure of data, e.g. pairwise distances.

Quantization

- Independent of sampling scheme.
- Coarse alphabets, e.g. $\{\pm 1\}$.
- Robust to hardware imperfections.
- Quantization error should decay super-linearly as function of measurements.

Wish List

Embedding

- Must be fast, e.g. convolutions, DFT, DCT.
- Approximately preserves structure of data, e.g. pairwise distances.

Quantization

- Independent of sampling scheme.
- Coarse alphabets, e.g. $\{\pm 1\}$.
- Robust to hardware imperfections.
- Quantization error should decay super-linearly as function of measurements.

Decoder

- Provable guarantees of accuracy with minimal measurements.
- Robust to model inaccuracies.
- Fast.

Applications

Any procedure that involves ε-nearest-neighbors searches:

- Classification: Classify a new object based on on the majority class of its neighbors.
- Regression: Assign value as the average or median of its neighbors.
- Data Retrieval: Find an object that resembles a particular query.
- Recommender Systems: Find a user who is most similar to a specific user.
- Clustering: k-means, ...

Johnson-Lindenstrauss Embeddings

Motivation: Random linear maps act as approximate isometries.

Lemma (Johnson, Lindenstrauss 1984)

Let $\mathcal{T} \subset \mathbb{R}^N$ be a finite set of points. For any

$$m \geq C \frac{\log(|\mathcal{T}|)}{\varepsilon^2},$$

there exists a (random) linear map $A : \mathbb{R}^N \to \mathbb{R}^m$ so that for any $x, y \in \mathcal{T}$,

$$\Big| \|Ax - Ay\|_2 - \|x - y\|_2 \Big| \leq \varepsilon \|x - y\|_2.$$

Quantized JL Embeddings

Idea: Use JL-embedding A and quantize each x to $\text{sign}(Ax)$.

Quantized JL Embeddings

Idea: Use JL-embedding A and quantize each x to $\text{sign}(Ax)$.

Lemma (Jacques et al 2011)

Let $\mathcal{T} \subset S^{N-1}$ be a finite set of points. For any

$$m \geq C' \frac{\log(|\mathcal{T}|)}{\varepsilon^2},$$

there exists a (random) linear map $A : \mathbb{R}^N \to \mathbb{R}^m$ so that for any $x, y \in \mathcal{T}$,

$$\Big| \|\text{sign}(Ax) - \text{sign}(Ay)\|_H - \|x - y\|_{S^{N-1}} \Big| \leq \varepsilon,$$

where $\|\cdot\|_H, \|\cdot\|_{S^{N-1}}$ are the normalized Hamming and geodesic distance, resp.

Quantized JL Embeddings

Idea: Use JL-embedding A and quantize each x to $\operatorname{sign}(Ax)$.

Lemma (Jacques et al 2011)

Let $\mathcal{T} \subset S^{N-1}$ be a finite set of points. For any

$$m \geq C' \frac{\log(|\mathcal{T}|)}{\varepsilon^2},$$

there exists a (random) linear map $A : \mathbb{R}^N \to \mathbb{R}^m$ so that for any $x, y \in \mathcal{T}$,

$$\Big| \| \operatorname{sign}(Ax) - \operatorname{sign}(Ay) \|_H - \|x - y\|_{S^{N-1}} \Big| \leq \varepsilon,$$

where $\|\cdot\|_H, \|\cdot\|_{S^{N-1}}$ are the normalized Hamming and geodesic distance, resp.

Remark: Up to constants, no extra price paid between JL and quantized JL embeddings!

Compressed Sensing Crash Course

Moving from finite sets to infinite sets requires more nuanced signal models.

- **Goal:** Recover $x \in \mathbb{R}^N$ from $y = Ax + \eta \in \mathbb{R}^m$, $m \ll N$, $\|\eta\|_2 \leq \varepsilon$ using structural priors on x, e.g. sparsity.
  - **Remark:** Signal acquisition via A: compressing while acquiring.
  - **Remark:** JL matrices often make good CS matrices.
- **Standard Solution**: Solve

$$x^\sharp := \arg\min \|z\|_1 \text{ s.t. } \|Az - y\|_2 \leq \varepsilon.$$

Theorem ((Candés et al 2006), (Cai et al 2014), ...)

If $A \in \mathbb{R}^{m \times N}$ satisfies $(2k, \alpha) - RIP$ with $\alpha \leq \frac{1}{\sqrt{2}}$ then

$$\|x^\sharp - x\|_2 \leq C_1 \varepsilon + C_2 \frac{\sigma_k(x)_1}{\sqrt{k}}, \quad \sigma_k(x)_1 = \min_{y \text{ } k\text{-sparse}} \|x - y\|_1.$$

Quantized Compressed Sensing

- As before, but now given $q = \mathcal{Q}(Ax) \in \mathcal{A}^m \subset \mathbb{R}^m$, $\mathcal{A}$ discrete.
    - Extreme case $\mathcal{A} = \{\pm 1\}$.
- Designing $\mathcal{Q}, \mathcal{A}$ is crucial to the analysis of problem.
- Much work has been done on recovering sparse vectors from quantized measurements, e.g.
    - P. Boufounos, R. Baraniuk "One-Bit Compressed Sensing," 2008.
    - S. Güntürk, A. Powell, R. Saab, O. Yilmaz "Sobolev Duals," 2010.
    - Y. Plan, R. Vershynin "Robust 1-bit Compressed Sensing," 2012.
    - L. Jacques, P. Boufounos, et al "Binary Stable Embeddings," 2015.
    - R. Saab, T. Huynh "Fast Binary Embeddings," 2018.

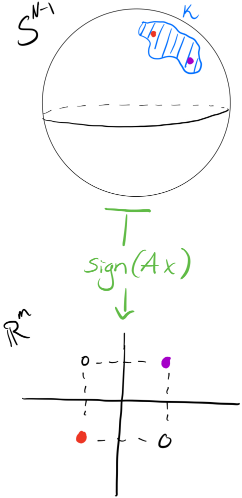
Quantized Embeddings of Infinite Sets

Few results on more general signal models, e.g.

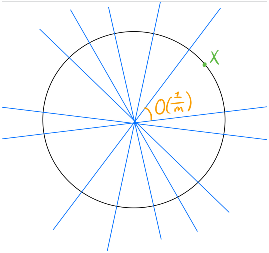
- V. Cambereri, L. Jacques "Time for Dithering," 2017.
- R. Vershynin and Y. Plan, "Robust 1-bit Compressed Sensing," 2013.
- M. Iwen, F. Krahmer et al "One-Bit Compressed Sensing on Manifolds," 2018.

These works admit at least one of the following shortcomings:

- Slow error decay as a function of m.
- Assumes you have parametrization of manifold.
- Limits model to be sub-manifold of S^{N-1}.
- Gaussian (read: slow) measurements.



Shortcomings of MSQ



- All magnitude information is lost: $\text{sign}(Ax) = \text{sign}(A\frac{x}{\|x\|})$.
- MSQ quantization (i.e. $y = \text{sign}(Ax)$) error cannot decay faster than $O(m^{-1})$ in frame setting [Goyal et al 1998].
- In sparse vector recovery [Romberg et al 2015] reconstruction error does not decay with m.

Our Set-Up

- **Signal Model:** (Unknown) d-manifold $K \subset B_2^N$.
- **Measurements:** $A \in \mathbb{R}^{m \times N}$ from sub-Gaussian ensemble, PCE, or BOE (rows selected uniformly with replacement). $D_\epsilon \in \mathbb{R}^{N \times N}$ diagonal of Rademacher r.v.'s independent of A.
    - **PCE**: Partial Circulant Ensemble. Appears in channel estimation, radar.
    - **BOE**: Bounded Orthonormal Ensemble, e.g. DFT, DCT. Appears in fMRI.
- **Quantization:** One-bit $q = \mathcal{Q}_{\Sigma\Delta}^{(r)}(AD_\epsilon x) =: \mathcal{Q}_{\Sigma\Delta}^{(r)}(\Phi x)$, (more later).
- **Approximatation of K:** *Geometric Multi-Resolution Analysis* (GMRA).
    - More later.

One-Bit Noise-Shaping Quantization

- Leverages correlations between measurements to minimize quantization error.

- For a given $r > 0$ and filter h with $|\operatorname{supp}(h)| = r$, define

$$q_i = \operatorname{sign}\left((h * u)_{i-1} + y_i\right),$$
$$u_i = (h * u)_{i-1} + y_i - q_i.$$

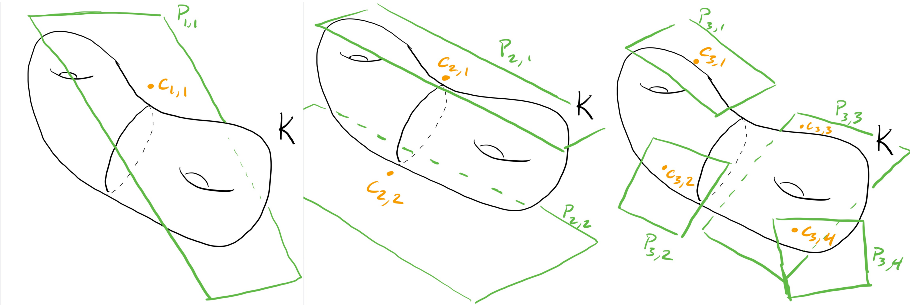
- Must choose h so that whenever $\|y\|_\infty < 1$, $\|u\|_\infty$ bounded by constant depending only on r.

Perks of Noise Shaping

- In frame and compressed sensing context, quantization error decays like $O(m^{-r})$ or $O(2^{-c'm})$.
  - P. Deift, S. Güntürk, F. Krahmer, “Exponentially Accurate One-Bit $\Sigma\Delta$” 2010.
  - R. Saab, R. Wang, O. Yilmaz, “Quantization of Compressive Samples” 2015.
  - E. Chou, S. Güntürk "Distributed Noise Shaping," 2016.
- In the above contexts, norm information is preserved.
- Certain instances of noise-shaping (e.g. $\Sigma\Delta$) are provably robust to hardware imperfections in machine arithmetic [Daubechies, Devore 2003]

GMRA

Roughly speaking, a GMRA is a sequence of affine approximations with a dyadic (tree) structure [Allard, Chen, Maggioni 2012].



GMRA at scale 0 GMRA at scale 1 GMRA at scale 2

Approximate Binary Embedding

Theorem (R. Saab, T. Huynh, 2018)

There exists $\widetilde{V} = p^{-1/2} \cdot I_{p\times p} \otimes \frac{v^T}{\|v\|_2} \in \mathbb{R}^{p\times m}$ *such that the following holds:*

let $K \subset B_1^N$, $m \geq p \geq C_1 \alpha^{-2} \log^4(N) \max\{1, \frac{w^2(K)}{\mathrm{rad}^2(K)}\}$, *and let* $f(x) = \mathcal{Q}_{\Sigma\Delta}^{(r)}(\Phi x)$. *Then with high probability*

$$\Big| \|\widetilde{V}(f(x) - f(y))\|_2 - \|x - y\|_2 \Big|$$

$$\leq \underbrace{\max\{\alpha, \sqrt{\alpha}\}\mathrm{rad}(K)}_{\text{manifold complexity}} + \underbrace{C_2 \left(\frac{m}{p}\right)^{-r+1/2}}_{\text{quantization error}}$$

for all $x, y \in K$.

Our Algorithm

GIVEN: $\Phi = AD_\varepsilon \in \mathbb{R}^{m\times N}$, $q := \mathcal{Q}(\Phi x)$, GMRA of K:

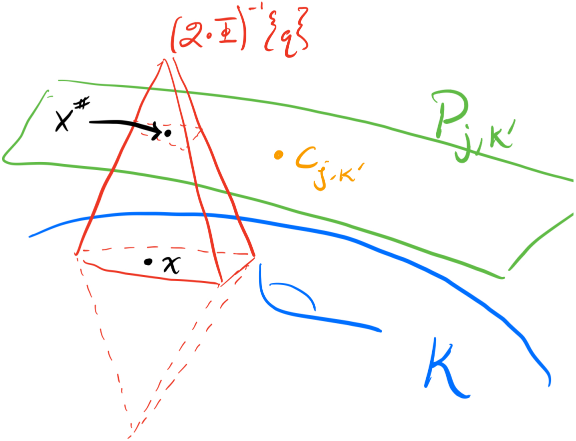
- **Step 1:** Find a center in the GMRA which quantizes to a bit-string close to q.

$$c_{j,k'} \in \operatorname*{arg\,min}_{c_{j,k}\in\mathcal{C}_j} \|\widetilde{V}(\mathcal{Q}(\Phi c_{j,k}) - q)\|_2.$$

- **Step 2:** Find the point in the GMRA closest to the quantization cell containing x.

$$x^\sharp = \operatorname*{arg\,min}_{z\in\mathbb{R}^N} \left\|\widetilde{V}\left(\Phi z - q\right)\right\|_2$$
$$\text{s.t. } z = P_{j,k'}(z),\ \|z\|_2 \leq 1.$$

Our Algorithm in Pictures



Our Result

Theorem (Iwen, L., Nelson, Saab 2019)

Let $S = K \cup$ GMRA at scale j. Suppose that

$$m \geq p \geq C \log^4(N) \frac{\max\{1, w^2(S)\mathrm{rad}^{-2}(S)\}}{\alpha^2},$$

and define $\lambda := m/p$. Then with high probability the following event occurs uniformly for all $x \in K$: the solution $x^\sharp$ of the main algorithm satisfies

$$\|x^\sharp - x\|_2 \lesssim_r \underbrace{C_x 2^{-j}}_{\text{GMRA Error}} + \underbrace{\max\{\sqrt{\alpha}, \alpha\}\mathrm{rad}(S)}_{\text{Manifold complexity.}} + \underbrace{\lambda^{-r+1/2}}_{\text{Quantization error}}.$$

Important Proof Ideas

- GMRA approximation incrementally improves with scale parameter.

  $\implies$ there's an affine plane approximating K nearby x.

- $V \circ \mathcal{Q} \circ \Phi$ approximate isometric embedding of $S = K \cup GMRA$ from $(\mathbb{R}^N, \ell_2)$ to $(\{\pm 1\}^p, \ell_2)$.

  $\implies$ Step 1 objective function $\|\widetilde{V}(\mathcal{Q}(\Phi c_{j,k}) - q)\|_2 \approx \|c_{j,k} - x\|_2$.

- $V \circ \Phi$ approximate isometric embedding of $S = K \cup GMRA$ from $(\mathbb{R}^N, \ell_2)$ to $(\mathbb{R}^p, \ell_2)$.

  $\implies$ Step 2 objective $\left\|\widetilde{V}\left(\Phi z - q\right)\right\|_2 \approx \|z - x\|_2 + \text{(small perturbation)}.$

Numerics

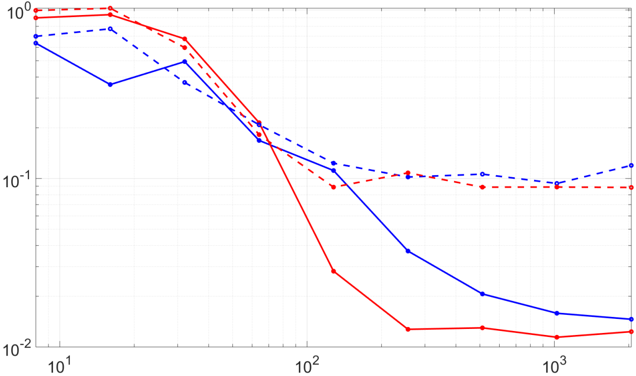


Figure: Log-scale error as a function of $\lambda = m/p$. Experiments for $S^2 \hookrightarrow \mathbb{R}^{20}$. Solid lines are GMRA refinement level $j = 12$; dashed lines to $j = 6$. Blue and red plots represent $r = 2, 4$ (resp.)

Concluding Remarks

- For approximately the same price (embedding dimension) as "analog" JL-embeddings, one can also find quantized JL-embeddings.
- The choice of encoder, particularly the quantizer, dramatically impacts quantization error decay of decoder.
- As in the frame/CS setting, noise-shaping quantizers exhibit same rapid quantization error decay in the (approximate) manifold model.

Appendix: Noise Shaping

Noise shaping quantizers with alphabet $\mathcal{A}$ and scalar quantizer $\mathcal{Q}(z) = \arg\min_{q \in \mathcal{A}} |q - z|$ update q, u via

$$\begin{aligned} q_i &= \mathcal{Q}(\rho(u_{i-r}, ..., u_{i-1}, y_i)), \\ y - q &= Hu, \end{aligned}$$

where H is lower-triangular (causality) and ρ is chosen so that $\|y\|_1 < C_1 \implies \|u\|_\infty < C_2$

Appendix: Noise Shaping

$\Sigma\Delta$: for $r > 0$,

$$H = \left(\begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & & \ddots & \\ & & -1 & 1 \end{bmatrix} \right)^r$$

[Daubechies, Devore 2003], [Güntürk, 2003], [Benedetto et al 2005]

Distributed Noise Shaping: for $\beta > 1$,

$$H_\beta = \begin{bmatrix} 1 & & & \\ -\beta & 1 & & \\ & & \ddots & \\ & & -\beta & 1 \end{bmatrix} \in \mathbb{R}^{m/p \times m/p},$$

$$H = \begin{bmatrix} H_\beta & & & \\ & H_\beta & & \\ & & \ddots & \\ & & & H_\beta \end{bmatrix} \in \mathbb{R}^{m \times m}$$

[Chou, Güntürk 2016]

Appendix: GMRA

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

$$\mathcal{P}_j = \left\{ P_{j,k} \colon \mathbb{R}^N \to \mathbb{R}^N : \; k \in [K_j] \right\}$$

with the following properties:

- **Affine Projections**. Every $P_{j,k}$ is an orthogonal projection onto some d-dimensional affine space which contains the center $c_{j,k}$.
- **Dyadic Structure**. The number of centers at each level is bounded by $|\mathcal{C}_j| = K_j \leq C_C 2^{dj}$ for an absolute constant $C_C \geq 1$. Moreover, there exist $C_1 > 0$, $C_2 \in (0, 1]$ such that
  - $K_j \leq K_{j+1}$ for all $j \in [J-1]$,
  - $\|c_{j,k_1} - c_{j,k_2}\|_2 > C_1 2^{-j}$ for all $j \in [J]$, $k_1 \neq k_2 \in [K_j]$,
  - For each $j \in [J] \setminus \{0\}$ there exists a parent function $p_j \colon [K_j] \to [K_{j-1}]$ with

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

......

Appendix: GMRA

- **Multiscale Approximation**. The projectors in $\mathcal{P}_j$ approximate K in the following sense:
  - There exists $j_0 \in [J-1]$ such that $c_{j,k} \in \mathrm{tube}_{C_1 2^{-j-2}}(K)$ for all $j \geq j_0$ and $k \in [K_j]$.
  - For each $j \in [J]$ and $z \in \mathbb{R}^N$, let

$$c_{j,k_j(z)} \in \underset{c_{j,k} \in \mathcal{C}_j}{\arg\min} \|z - c_{j,k}\|_2.$$

  Then for each $z \in K$ there exist $C_z, \widetilde{C}_z > 0$ so that $\|z - P_{j,k_j(z)}z\|_2 \leq C_z 2^{-2j}$ for all $j \in [J]$ and

$$\|z - P_{j,k'}z\|_2 \leq \widetilde{C}_z 2^{-j}$$

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

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