Quantized compressed sensing and quasi-isometric embeddings

Laurent Jacques
UCLouvain, Belgium

May 18th, 2015

Outline

I. Ultra brief intro to CS facts
II. Quantization context
III. Initial approaches for quantized CS
IV. Toward consistency in QCS
V. 1-bit CS and Binary ε-Stable Embedding (BεSE)
VI. Quantized Embeddings

Take-away messages & open questions

I. Ultra brief intro to CS

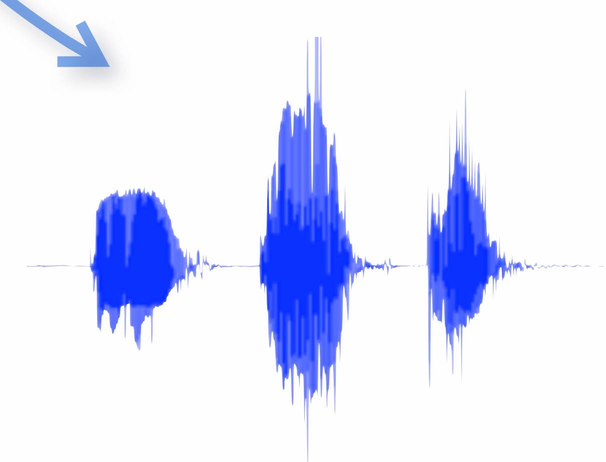
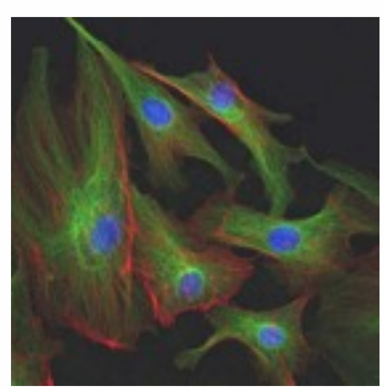
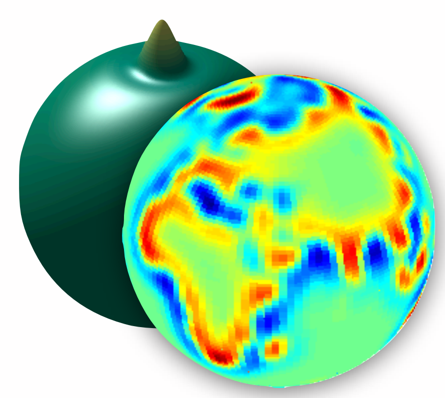
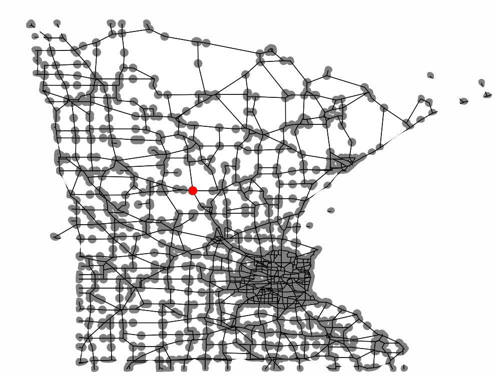
Compressed Sensing...

… *in a nutshell:*

Generalize Dirac/Nyquist sampling:

1°) ask *few* (**linear**) *questions*
about your ***informative*** signal

2°) and recover it *differently* (**non-linearly**)”

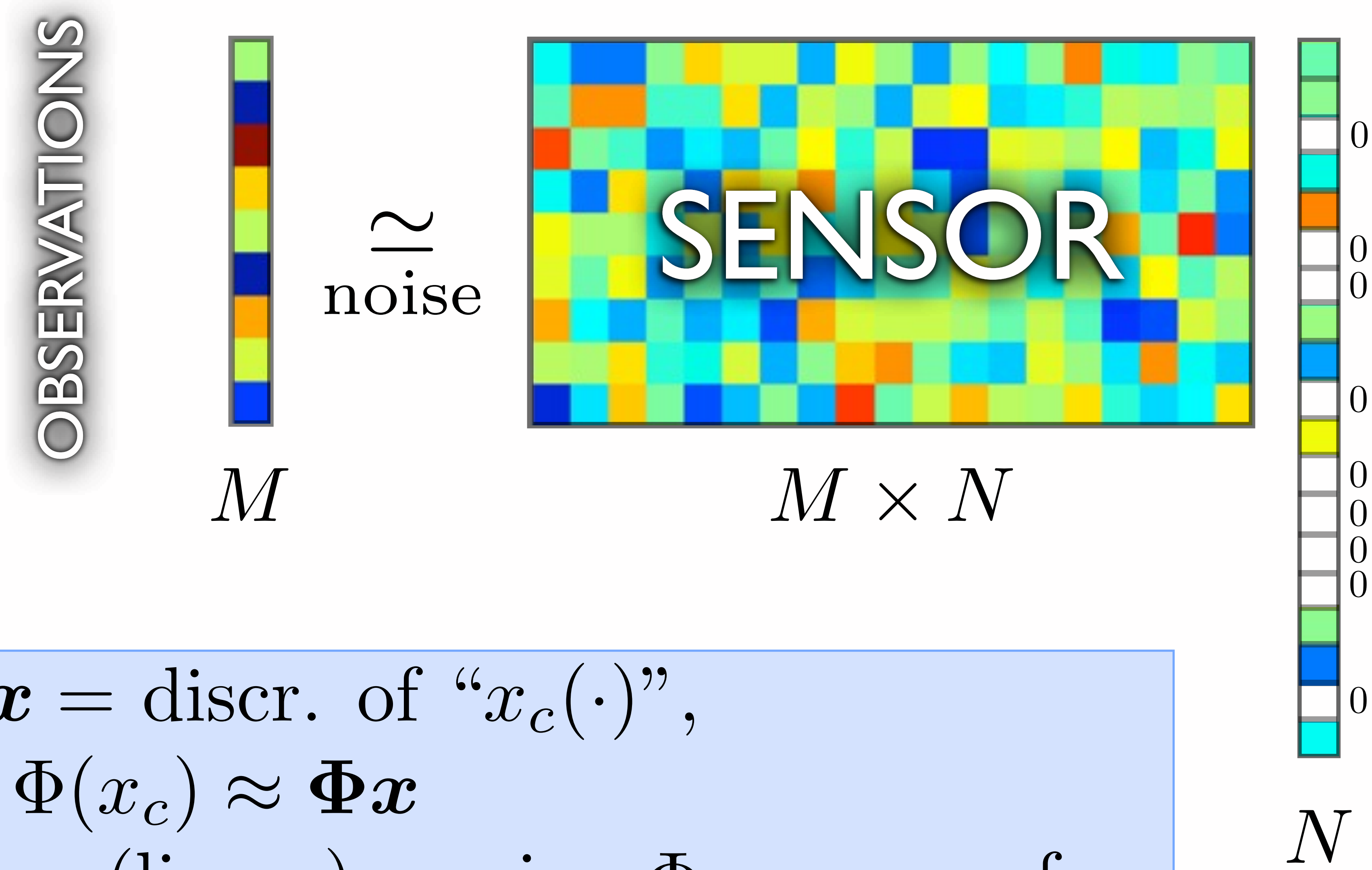





e.g., sparse, structured, low-rank, ...

1st, CS $\ni$ Generalized Linear Sensing!

M questions — Sensing method — Signal

$\boldsymbol{y}$ — $\boldsymbol{\Phi}$ — $\boldsymbol{x}$



Sparsity Prior ($\Psi = \mathrm{Id}$)

A signal in this discrete world

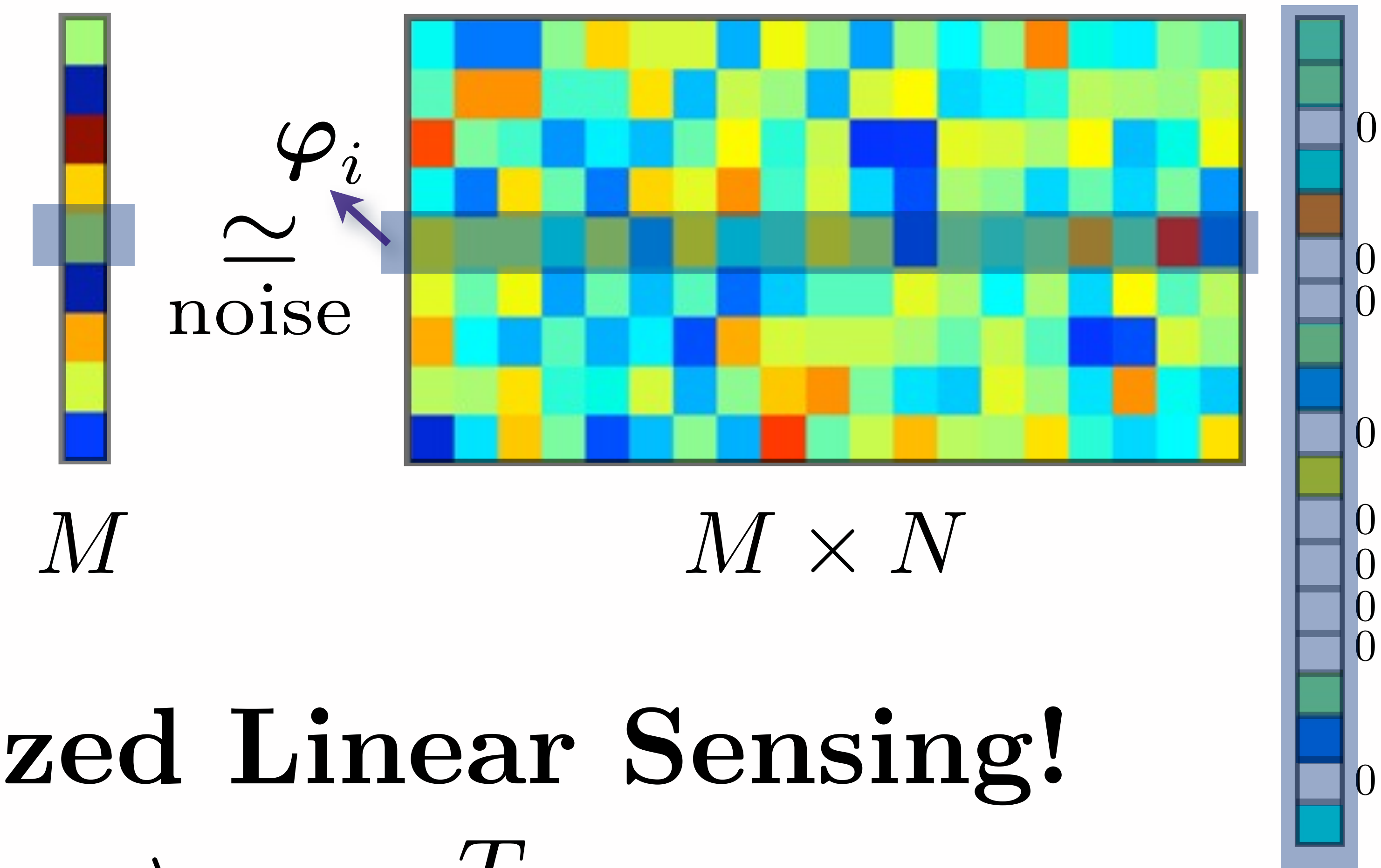
Caveat: $\boldsymbol{x}$ = discr. of "$x_c(\cdot)$",
and $\boldsymbol{y} = \Phi(x_c) \approx \boldsymbol{\Phi x}$
given some (linear) sensing Φ process of x_c

1st, CS $\ni$ Generalized Linear Sensing!

M questions | Sensing method | Signal

$\boldsymbol{y}$ $\quad$ $\boldsymbol{\Phi}$ $\quad$ $\boldsymbol{x}$

y_i $\underset{\text{noise}}{\simeq}$ $\boldsymbol{\varphi}_i$



M $\quad$ $M \times N$ $\quad$ N

Sparsity
Prior
$(\Psi = \mathrm{Id})$

A signal
in this
discrete
world

Generalized Linear Sensing!

$$y_i \simeq \langle \boldsymbol{\varphi}_i, \boldsymbol{x} \rangle = \boldsymbol{\varphi}_i^T \boldsymbol{x}$$

$$1 \leq i \leq M$$

e.g., to be realized
optically/analogically

2nd, CS $\ni$ Non-linear reconstruction!

<u>Possible reconstruction:</u> (others exist)

(Basis Pursuit DeNoise)

[Chen, Donoho, Saunders, 1998]

$$\hat{\boldsymbol{x}} = \operatorname*{arg\,min}_{\boldsymbol{u} \in \mathbb{R}^N} \|\boldsymbol{u}\|_1 \text{ s.t. } \|\boldsymbol{y} - \boldsymbol{\Phi u}\| \leq \epsilon$$

Sparsity promotion: $\|\boldsymbol{u}\|_1 = \sum_j |u_j|$

Level of "noise"

2nd, CS $\ni$ Non-linear reconstruction!

BPDN instance optimality:

If $\frac{1}{\sqrt{M}}\boldsymbol{\Phi}$ respects the Restricted Isometry Property (RIP)

$$\boxed{(1-\delta)\|\boldsymbol{u}\|^2 \leq \tfrac{1}{M}\|\boldsymbol{\Phi u}\|^2 \leq (1+\delta)\|\boldsymbol{u}\|^2}$$

for all $\boldsymbol{u} \in \Sigma_{2K} := \{\boldsymbol{u} : \|\boldsymbol{u}\|_0 := |\mathrm{supp}\,\boldsymbol{u}| \leq 2K\}$

Then, if $\delta < \sqrt{2} - 1$ [Candès, 09],

Robustness: *vs* sparse deviation + noise.

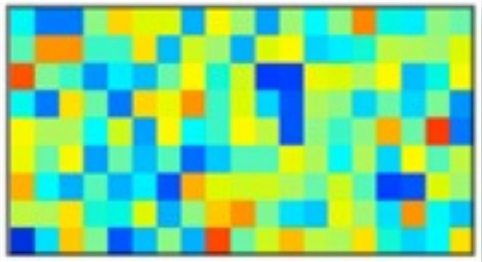
$$\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| \ \lesssim \ \frac{1}{\sqrt{K}}\|\boldsymbol{x} - \boldsymbol{x}_K\|_1 + \frac{\epsilon}{\sqrt{M}}$$

(with $f \lesssim g \ \equiv \ \exists c > 0 : f \leqslant c\,g$)

2nd, CS $\ni$ Non-linear reconstruction!

Matrices with RIP?

$\boldsymbol{\Phi} \in \mathbb{R}^{M\times N}$, with $\Phi_{ij} \sim_{\rm iid} \mathcal{N}(0,1)$ and $M \gtrsim K \log N/K$.



but also:

- Random sub-Gaussian ensembles (e.g., Bernoulli);
- random Fourier/Hadamard ensembles (structured sensing);
- random convolutions, spread-spectrum;
- ...

(see, *e.g.*, "A Mathematical Introduction to Compressive Sensing", Rauhut, Foucart, Springer, 2013)

II. Quantization context

(Restricted to scalar quantization)

Caveat: not covered here:

- Sigma-Delta quantization for CS

  (see, *e.g.*, Kramer, Saab, Güntürk, Powell, Ward, ...)
- Vector quantization

  (see, *e.g.*, Goyal, Nguyen, Sun, ...)
- Universal quantization (periodic)

  (see, *e.g.*, Boufounos, Rane, ...)

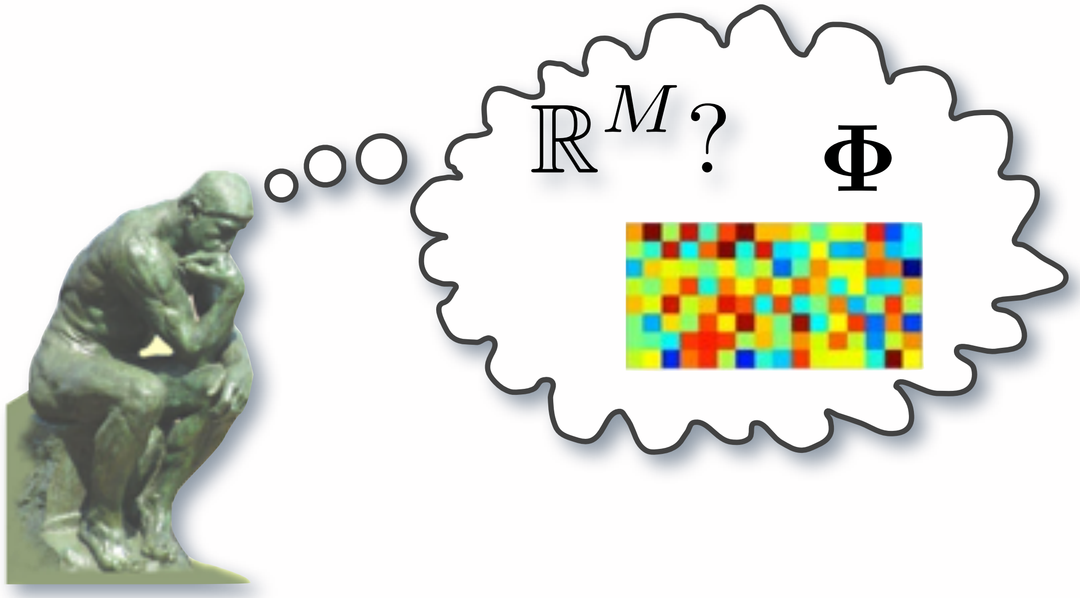
Compressive Sampling and Quantization

Compressed sensing theory says:

"Linearly sample a signal

at a rate function of

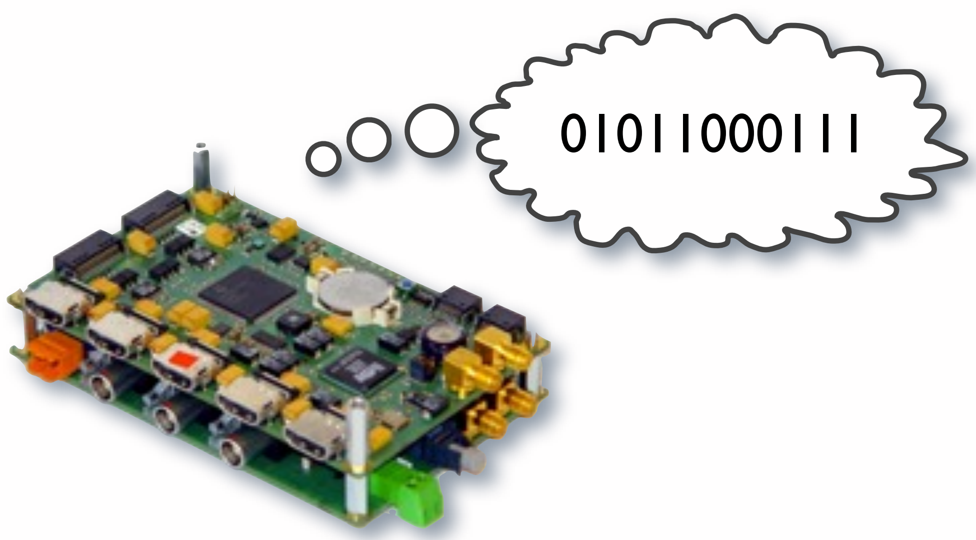
its intrinsic dimensionality"



Information theory and sensor designer say:

"Okay, but I need to quantize/digitize my measurements!"

(e.g., in ADC)



Integration?

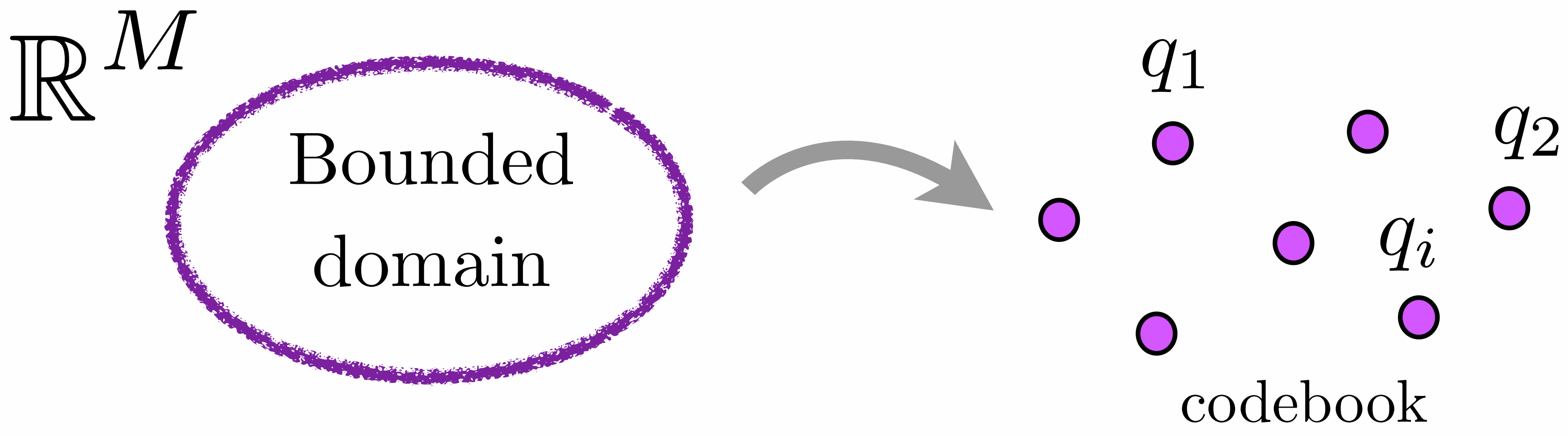
QCS theory?

Theoretical Bounds

What is quantization?

- Generality:

  Intuitively: "*Quantization maps a bounded continuous domain to a set of finite elements (or codebook)*"



$$\mathcal{Q}[x] \in \{q_1, q_2, \cdots\}$$

- Oldest example: rounding off $\lfloor x \rfloor, \lceil x \rceil, \ldots \quad \mathbb{R} \to \mathbb{Z}$

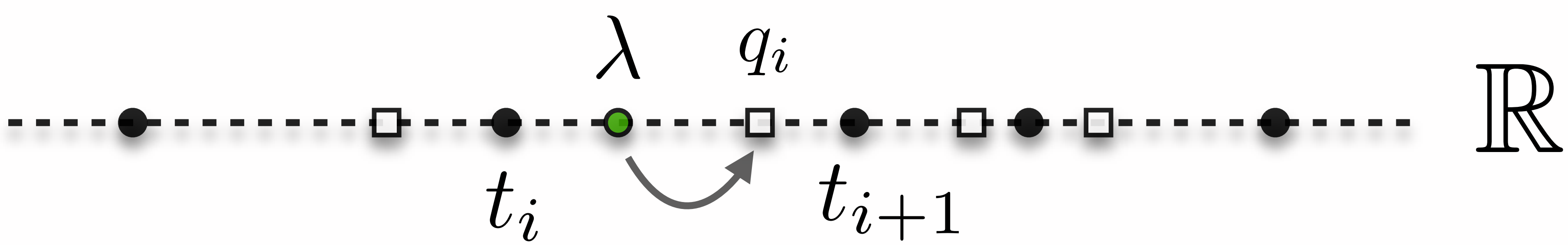
Scalar quantization

Pulse Code Modulation - PCM
Memoryless Scalar Quantization - MSQ

Applied on each component of M- dimensional vectors:

$$\mathcal{Q}(\lambda) = q_i$$

□ : Level $\Omega = \{q_i\}$ (or codebook) ● : Thresholds $\mathcal{T} = \{t_i\}$



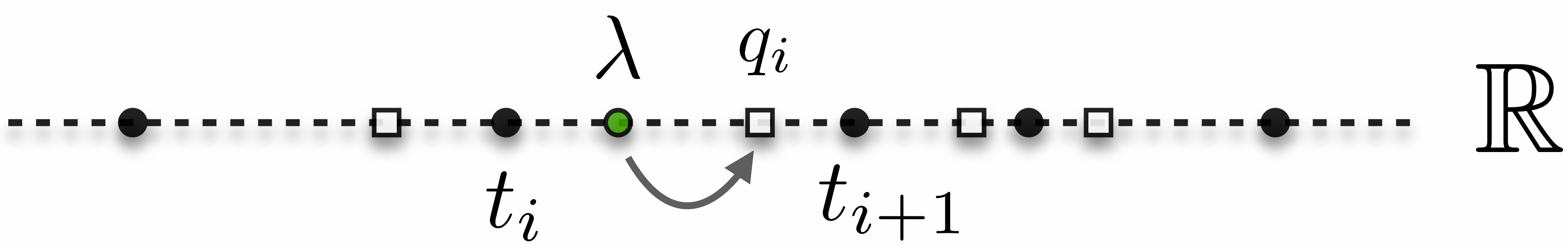
Scalar quantization

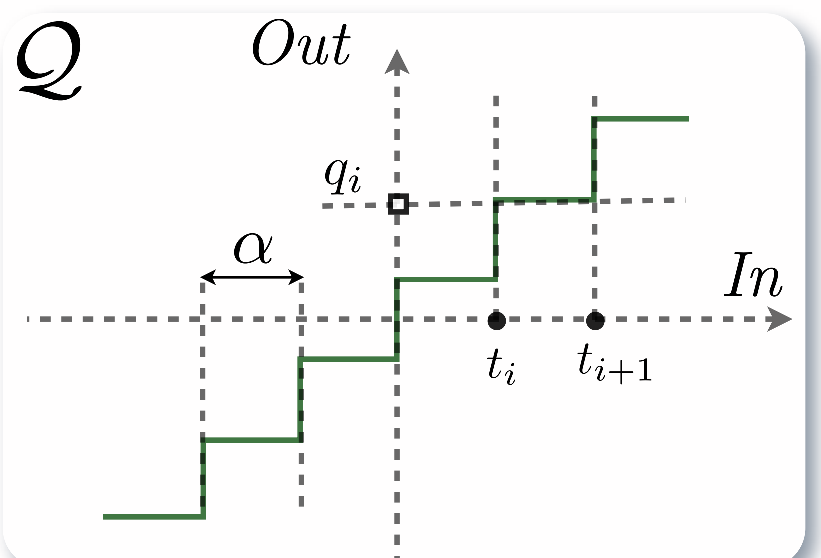
Pulse Code Modulation - PCM
Memoryless Scalar Quantization - MSQ

Applied on each component of M- dimensional vectors:

$$\mathcal{Q}(\lambda) = q_i$$

$\square$: Level $\Omega = \{q_i\}$ (or codebook) $\bullet$: Thresholds $\mathcal{T} = \{t_i\}$





Example: uniform, resolution α

$$q_k = (k + 1/2)\alpha$$

$$t_k = k\alpha$$

$$\mathcal{Q}(t) = \alpha(\lfloor \tfrac{t}{\alpha} \rfloor + \tfrac{1}{2})$$

... with possible non-uniform adaption (Lloyd-Max)

Quantizing Compressed Sensing?

With no additional noise:

e.g., basis pursuit,
greedy methods, ...

$$\boldsymbol{x} \xrightarrow{\mathbb{R}^N} \boxed{\mathsf{CS}} \xrightarrow[]{\overset{\mathbb{R}^M}{\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}}} \underset{}{\overset{\underset{\text{codebook}}{\boldsymbol{\Omega}}}{\boxed{\mathcal{Q}}}} \xrightarrow{\boldsymbol{q} = \mathcal{Q}[\boldsymbol{y}]} \boxed{\mathsf{Decoder}} \xrightarrow{\mathbb{R}^N} \hat{\boldsymbol{x}}$$

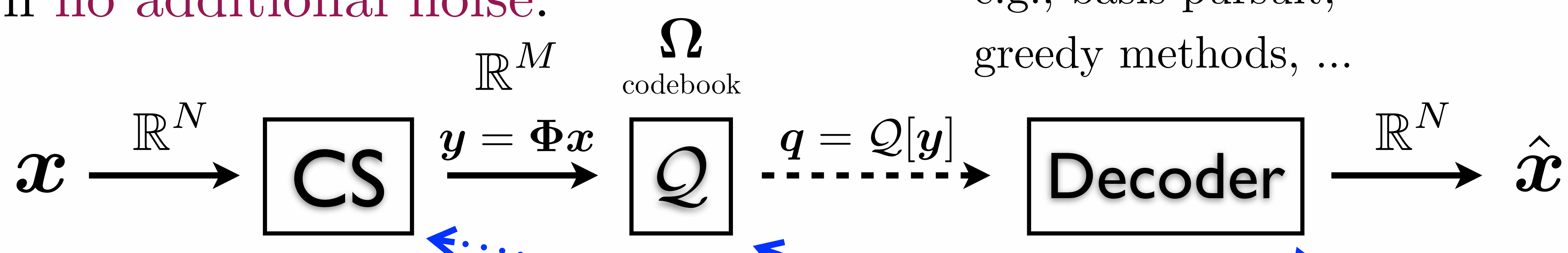
Finite codebook $\Rightarrow \hat{\boldsymbol{x}} \neq \boldsymbol{x}$

$i.e.$, impossibility to encode continuous domain
in a finite number of elements.

Quantizing Compressed Sensing?

With no additional noise:

e.g., basis pursuit,
greedy methods, ...



Finite codebook $\Rightarrow \hat{\boldsymbol{x}} \neq \boldsymbol{x}$

i.e., impossibility to encode continuous domain
in a finite number of elements.

Objective: Minimize $\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|$

given a certain number of:
bits, measurements, or bits/meas.

Where to act?
Change CS, Q or decoder?
Some of them? all?

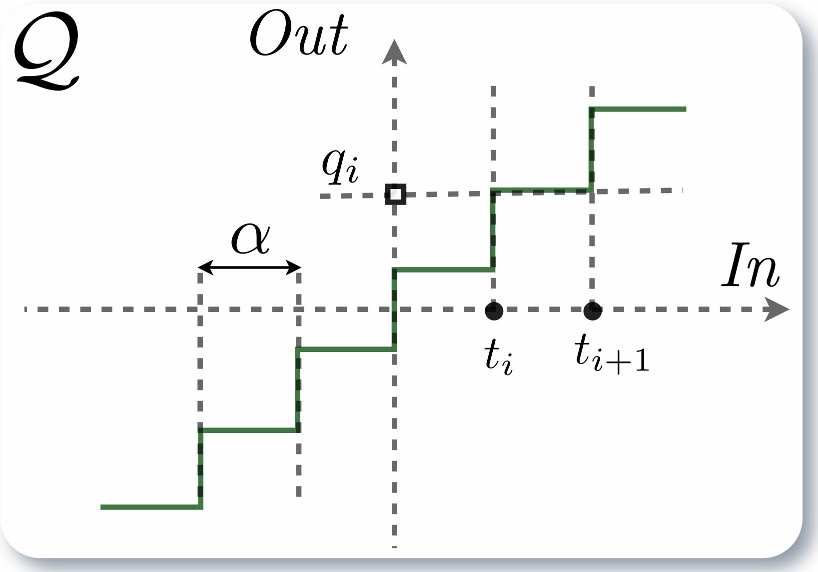
III. Initial Approaches for Quantized CS

Former solution (Candès, Tao, ...)

- (scalar) Quantization is like a noise

$$\boxed{\boldsymbol{q} \;=\; \mathcal{Q}\big[\boldsymbol{\Phi x}\big] = \boldsymbol{\Phi x} + \boldsymbol{n}}$$

quantization distortion



Former solution (Candès, Tao, ...)

- (scalar) Quantization is like a noise

$$\boxed{\boldsymbol{q} \;=\; \mathcal{Q}\big[\boldsymbol{\Phi x}\big] = \boldsymbol{\Phi x} + \boldsymbol{n}}$$

- **and** CS is robust (e.g., with *basis pursuit denoise*)

$$\hat{\boldsymbol{x}} = \operatorname*{argmin}_{\boldsymbol{u}\in\mathbb{R}^N} \|\boldsymbol{u}\|_1 \text{ s.t. } \|\boldsymbol{\Phi u} - \boldsymbol{q}\| \leqslant \epsilon \quad \text{(BPDN)}$$

$\ell_2 - \ell_1$ instance optimality:

If $\|\boldsymbol{n}\| \leqslant \epsilon$ and $\frac{1}{\sqrt{M}}\boldsymbol{\Phi}$ is RIP$(\delta, 2K)$ with $\delta \leqslant \sqrt{2} - 1$, then

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\| \;\lesssim\; \tfrac{\epsilon}{\sqrt{M}} + e_0(K),$$

with $e_0(K) = \|\boldsymbol{x} - \boldsymbol{x}_K\|_1/\sqrt{K}$.

Former solution (Candès, Tao, ...)

- (scalar) Quantization is like a noise

$$\boxed{\boldsymbol{q} \;=\; \mathcal{Q}[\boldsymbol{\Phi x}] = \boldsymbol{\Phi x} + \boldsymbol{n}}$$

- **and** CS is robust (e.g., with *basis pursuit denoise*)

$$\hat{\boldsymbol{x}} = \operatorname*{argmin}_{\boldsymbol{u}\in\mathbb{R}^N} \|\boldsymbol{u}\|_1 \text{ s.t. } \|\boldsymbol{\Phi u} - \boldsymbol{q}\| \leqslant \epsilon \quad \text{(BPDN)}$$

$\ell_2 - \ell_1$ instance optimality:

If $\|\boldsymbol{n}\| \leqslant \epsilon$ and $\frac{1}{\sqrt{M}}\boldsymbol{\Phi}$ is RIP$(\delta, 2K)$ with $\delta \leqslant \sqrt{2} - 1$, then

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\| \;\lesssim\; \tfrac{\epsilon}{\sqrt{M}} + e_0(K),$$

with $e_0(K) = \|\boldsymbol{x} - \boldsymbol{x}_K\|_1/\sqrt{K}$.

Deterministic: $\epsilon^2 \leq M\alpha^2/4$
Stochastic: $\epsilon^2 \leq M\alpha^2/12 + c\sqrt{M}$ (w.h.p)

Former solution (Candès, Tao, ...)

- From BPDN $\ell_2 - \ell_1$ instance optimality:

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\| \lesssim \alpha + e_0(K),$$

- Other reading :
  - B bits per measurements $\Rightarrow \alpha \propto 2^{-B}$

$$\Rightarrow \text{BPDN RMSE} \lesssim 2^{-B} + e_0(K)$$

when RIP holds, *i.e.*, for $M = O(K \log N/K)$

Former solution (Candès, Tao, ...)

- From BPDN $\ell_2 - \ell_1$ instance optimality:

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\| \lesssim \alpha + e_0(K),$$

- Other reading :
  - B bits per measurements $\Rightarrow \alpha \propto 2^{-B}$

$$\Rightarrow \text{BPDN RMSE} \lesssim 2^{-B} + e_0(K)$$

when RIP holds, *i.e.*, for $M = O(K \log N/K)$

But quantization error doesn't decay with M !?

Solution: be consistent!

IV. Toward consistency in Quantized CS

under High Resolution Assumption (HRA)

Consistent reconstructions in CS?

Issue: if $\hat{\boldsymbol{x}}$ solution of BPDN,

$$\hat{\boldsymbol{x}} = \operatorname*{argmin}_{\boldsymbol{u}\in\mathbb{R}^N} \|\boldsymbol{u}\|_1 \text{ s.t. } \|\boldsymbol{\Phi u} - \boldsymbol{q}\| \leqslant \epsilon \quad \text{(BPDN)}$$

(i) No Quantization Consistency (QC) !

$$\underbrace{\|\boldsymbol{\Phi}\hat{\boldsymbol{x}} - \mathcal{Q}[\boldsymbol{\Phi x}]\| \leqslant \epsilon_2}_{\text{(from BPDN constraint)}} \quad \not\Rightarrow \mathcal{Q}[\boldsymbol{\Phi}\hat{\boldsymbol{x}}] = Q[\boldsymbol{\Phi x}]$$

$$\Rightarrow \boxed{\mathcal{Q}[\boldsymbol{\Phi}\hat{\boldsymbol{x}}] \neq Q[\boldsymbol{\Phi x}]}$$

$\Rightarrow$ Sensing information is fully not exploited!

(ii) ℓ_2 constraint in BPDN

$\approx$ Gaussian distribution (MAP - cond. log. lik.)

But why looking for consistency?

First: Let T the support of $\boldsymbol{x} \in \mathbb{R}^N$, $r = M/K$, $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$, and $\boldsymbol{y} = \boldsymbol{\Phi x}$.

oversampling *rate*

Proposition (Goyal, Vetterli, Thao, 98) *If T is known (with $|T| = K$), the best decoder* Dec() *provides a $\hat{\boldsymbol{x}} = \mathrm{Dec}(\boldsymbol{y}, \boldsymbol{\Phi})$ such that:*

$$\mathrm{RMSE} = (\mathbb{E}\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2)^{1/2} \gtrsim \underbrace{r^{-1}}_{O(1/M)}\alpha,$$

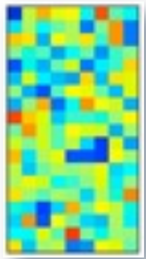
where $\mathbb{E}$ is wrt a probability measure on $\boldsymbol{x}_T$ in a bounded set $\mathcal{S} \subset \mathbb{R}^K$.

This bound is achieved, at least, for $\boldsymbol{\Phi}_T = \mathrm{DFT} \in \mathbb{R}^{M \times K}$, when Dec() *is* **consistent.**

V. K Goyal, M. Vetterli, N. T. Thao, "Quantized Overcomplete Expansions in R^N: Analysis, Synthesis, and Algorithms" , IEEE Tran. IT, 44(1), 1998

But why looking for consistency?

N

M 

<u>Second</u>,

If $\boldsymbol{\Phi} \in \mathbb{R}^{M\times N}$ is a (random) frame in $\mathbb{R}^N$ $(M \geqslant N)$ and $\boldsymbol{y} = \boldsymbol{\Phi x}$,

Then, for $\mathcal{Q}(\boldsymbol{y}) = \boldsymbol{y} + \boldsymbol{\xi}$ with $\xi_i \sim \mathcal{U}([-\tfrac{1}{2}\alpha, \tfrac{1}{2}\alpha])$,

(achievable with *dithering* or under HRA)

and $\hat{\boldsymbol{x}}$ **consistent**,

$$\big(\mathbb{E}_{\boldsymbol{\Phi},\boldsymbol{n}}\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2\big)^{1/2} \;\lesssim\; \big(\tfrac{M}{N}\big)^{-1}\alpha,$$

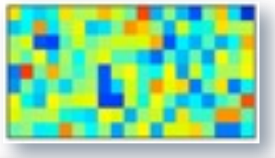
[Powell, Whitehouse, 2013]
(unit norm frame)

$$\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| \;\lesssim\; \big(\tfrac{M}{N}\big)^{-1}\alpha \cdot O(\log M, \log N, \log \eta),$$

with $\Pr \geqslant 1 - \eta$.

[LJ 2014]
(Gaussian frame)

or $\big(\tfrac{M}{K}\big)^{-1}$ if $\boldsymbol{x}$ is K-sparse with Gaussian sensing matrix.



In quest of consistency...

$\ell_2 \to \ell_\infty$

- Modify BPDN [W. Dai, O. Milenkovic, 09]

$$\hat{\boldsymbol{x}} = \operatorname*{argmin}_{\boldsymbol{u}\in\mathbb{R}^N} \|\boldsymbol{u}\|_1 \text{ s.t. } \mathcal{Q}[\boldsymbol{\Phi}\boldsymbol{u}] = \boldsymbol{q}$$

$$\Leftrightarrow \|\boldsymbol{\Phi}\boldsymbol{u} - \boldsymbol{q}\|_\infty \leq \alpha/2$$

+ modified greedy algo: *"subspace pursuit"*

W. Dai, H. V. Pham, and O. Milenkovic, "Quantized Compressive Sensing", preprint, 2009

In quest of consistency... $\ell_2 \to \ell_\infty$

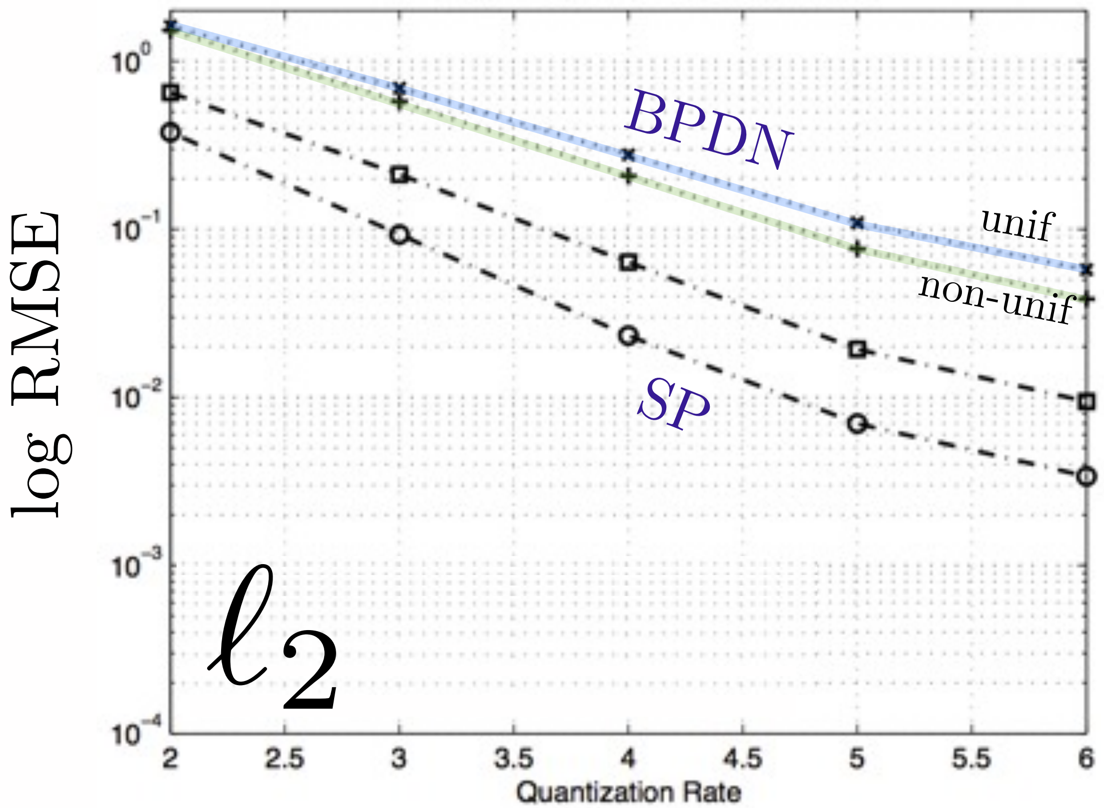
- Modify BPDN [W. Dai, O. Milenkovic, 09]

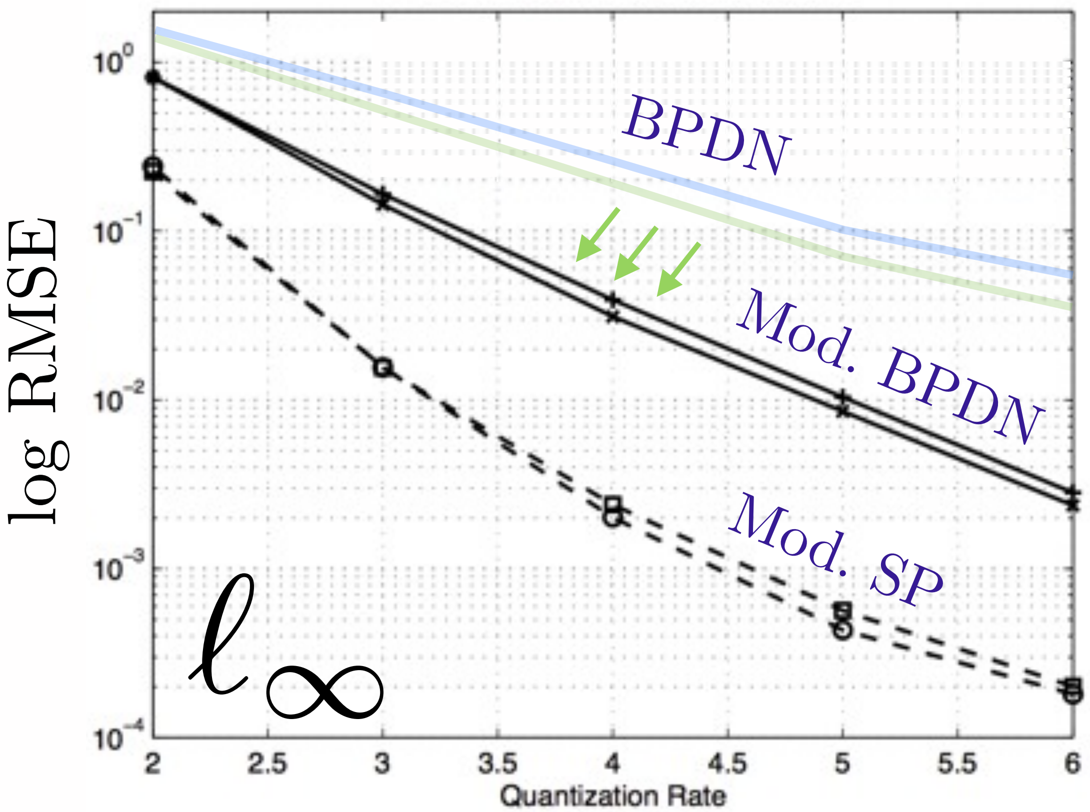
$$\hat{\boldsymbol{x}} = \operatorname*{argmin}_{\boldsymbol{u}\in\mathbb{R}^N} \|\boldsymbol{u}\|_1 \text{ s.t. } \mathcal{Q}[\boldsymbol{\Phi u}] = \boldsymbol{q}$$

$$\Leftrightarrow \|\boldsymbol{\Phi u} - \boldsymbol{q}\|_\infty \leq \alpha/2$$

+ modified greedy algo: *"subspace pursuit"*

Simulations: $M = 128, N = 256, K = 6, 1000$ trials $\Rightarrow \lambda \simeq 20$



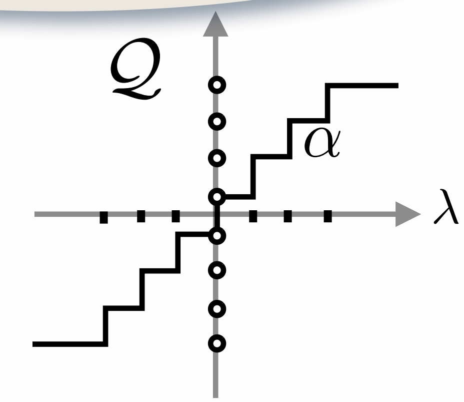


W. Dai, H. V. Pham, and O. Milenkovic, "Quantized Compressive Sensing", preprint, 2009

Dequantizing CS?

[LJ, Hammond, Fadili, 2009, 2011]

$\ell_2 \to \ell_p \ (p \geq 2)$



- Distortion model: $\|\boldsymbol{\Phi x} - \boldsymbol{q}\|_\infty \leq \alpha/2$

$$\boldsymbol{q} = \mathcal{Q}[\boldsymbol{\Phi x}] = \boldsymbol{\Phi x} + \boldsymbol{n}, \quad n_i \sim U(-\tfrac{\alpha}{2}, \tfrac{\alpha}{2})$$

- Reconstruction: Basis Pursuit DeQuantizer

$$\boxed{\hat{\boldsymbol{x}} \ = \ \underset{\boldsymbol{u} \in \mathbb{R}^N}{\arg\min} \|\boldsymbol{u}\|_1 \text{ s.t. } \|\boldsymbol{q} - \boldsymbol{\Phi u}\|_p \ \leq \epsilon_p}$$

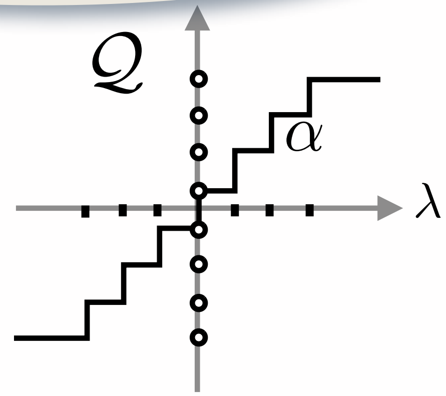
Towards $p = \infty$

Related to GGD MAP

Dequantizing CS?

[LJ, Hammond, Fadili, 2009, 2011]

$\ell_2 \to \ell_p \; (p \geq 2)$



- Distortion model: $\|\boldsymbol{\Phi x} - \boldsymbol{q}\|_\infty \leq \alpha/2$

$$\boldsymbol{q} = \mathcal{Q}[\boldsymbol{\Phi x}] = \boldsymbol{\Phi x} + \boldsymbol{n}, \quad n_i \sim U(-\tfrac{\alpha}{2}, \tfrac{\alpha}{2})$$

- Reconstruction: Basis Pursuit DeQuantizer

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{u} \in \mathbb{R}^N} \|\boldsymbol{u}\|_1 \text{ s.t. } \|\boldsymbol{q} - \boldsymbol{\Phi u}\|_p \leq \epsilon_p$$

BPDQ Stability?

Ok, if $\boldsymbol{\Phi}$ is RIP_p of order K, *i.e.*, (for $\mu_p \simeq M^{1/p}$)

$\exists\ \mu_p > 0,\ \delta \in (0,1),$

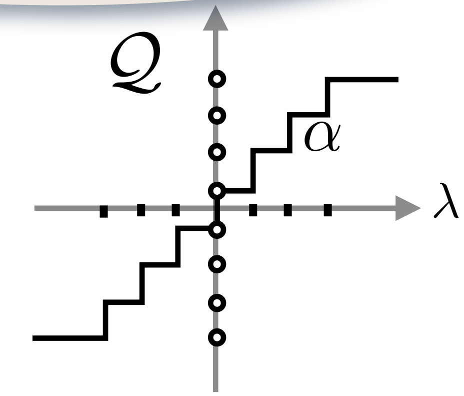
$$\sqrt{1-\delta}\,\|\boldsymbol{v}\|_2 \leqslant \frac{1}{\mu_p}\|\boldsymbol{\Phi v}\|_p \leqslant \sqrt{1+\delta}\,\|\boldsymbol{v}\|_2,$$

for all K sparse signals $\boldsymbol{v}$.

Dequantizing CS?

[LJ, Hammond, Fadili, 2009, 2011]

$\ell_2 \to \ell_p \ (p \geq 2)$



- Distortion model: $\|\boldsymbol{\Phi x} - \boldsymbol{q}\|_\infty \leq \alpha/2$

$$\boldsymbol{q} = \mathcal{Q}[\boldsymbol{\Phi x}] = \boldsymbol{\Phi x} + \boldsymbol{n}, \quad n_i \sim U(-\tfrac{\alpha}{2}, \tfrac{\alpha}{2})$$

- Reconstruction: Basis Pursuit DeQuantizer

$$\hat{\boldsymbol{x}} = \operatorname*{arg\,min}_{\boldsymbol{u} \in \mathbb{R}^N} \|\boldsymbol{u}\|_1 \text{ s.t. } \|\boldsymbol{q} - \boldsymbol{\Phi u}\|_p \leq \epsilon_p$$

BPDQ Stability?

Gain over BPDN (for tight $\epsilon_p(\alpha, M)$)

$$\Rightarrow \|\boldsymbol{x} - \hat{\boldsymbol{x}}\| = O(\epsilon_p/\mu_p)$$

$$\Rightarrow \|\boldsymbol{x} - \hat{\boldsymbol{x}}\| = O(\alpha/\sqrt{p+1})$$

But no free lunch: for $\boldsymbol{\Phi}$ Gaussian

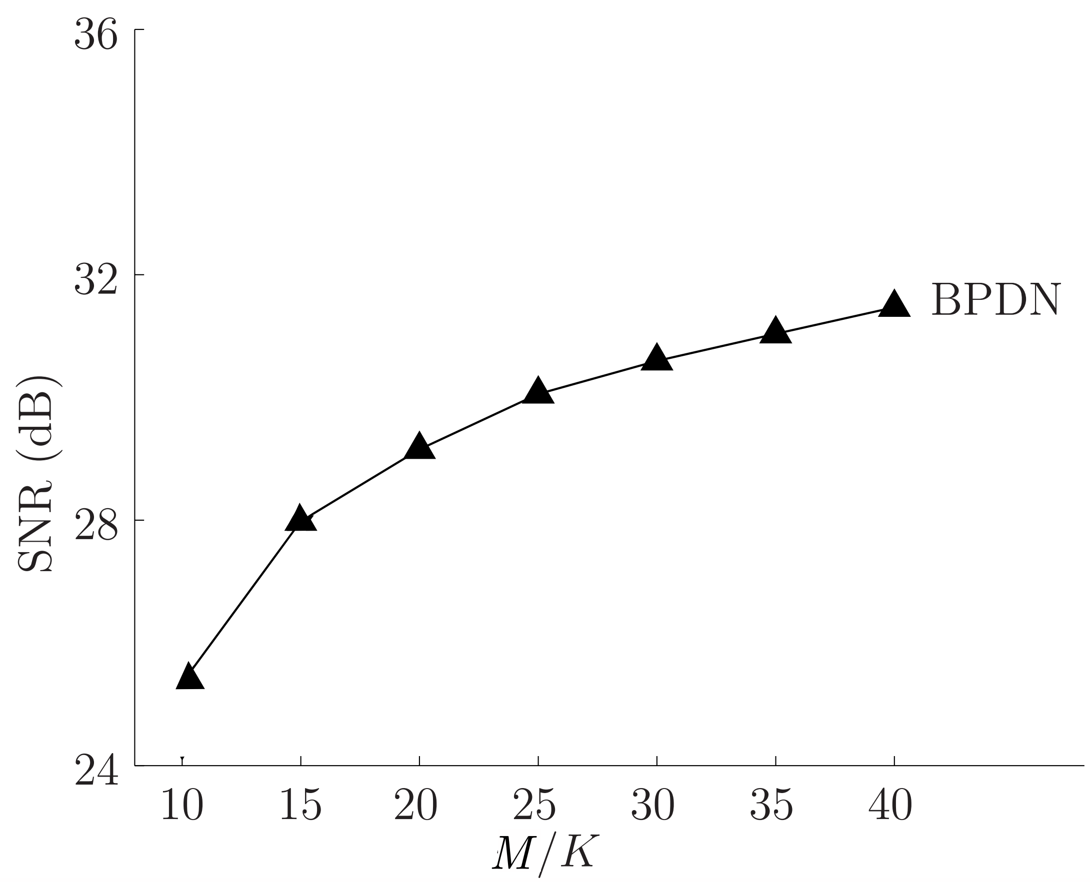
$$M = O\big(\,(K \log N/K\,)^{p/2}\,\big)$$

$\Longrightarrow$ Another reading: limited range of valid p for a given M (and K)!

(note: RIP$_p$ not required anymore, but diff. error bound. S. Dirksen et al, '15)

Dequantizing CS?

[LJ, Hammond, Fadili, 2009, 2011]

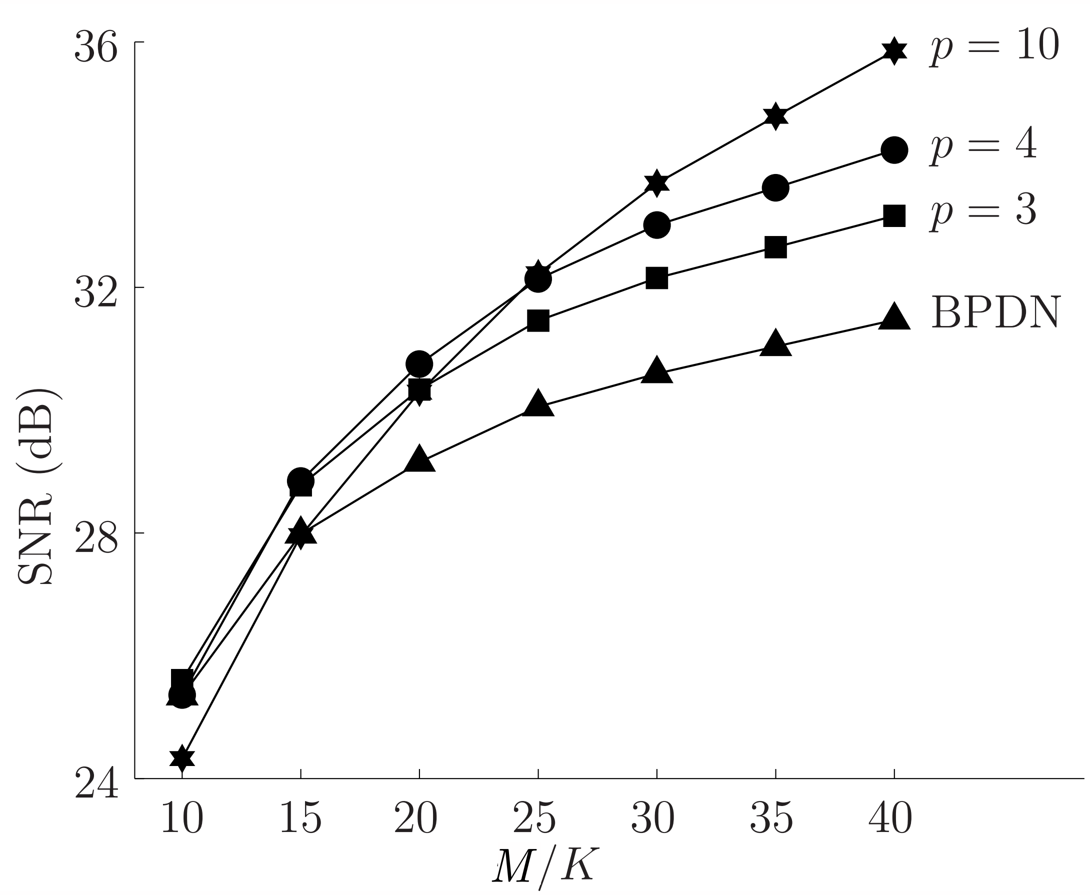


- N=1024, K=16, Gaussian $\mathbf{\Phi}$
- 500 K-sparse (canonical basis)
- Non-zero components follow $\mathcal{N}(0,1)$
- Quantiz. bin width $\alpha = \|\mathbf{\Phi}\boldsymbol{x}\|_\infty/40$

LJ, D. Hammond, J. Fadili "Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine." Information Theory, IEEE Transactions on, 57(1), 559-571.

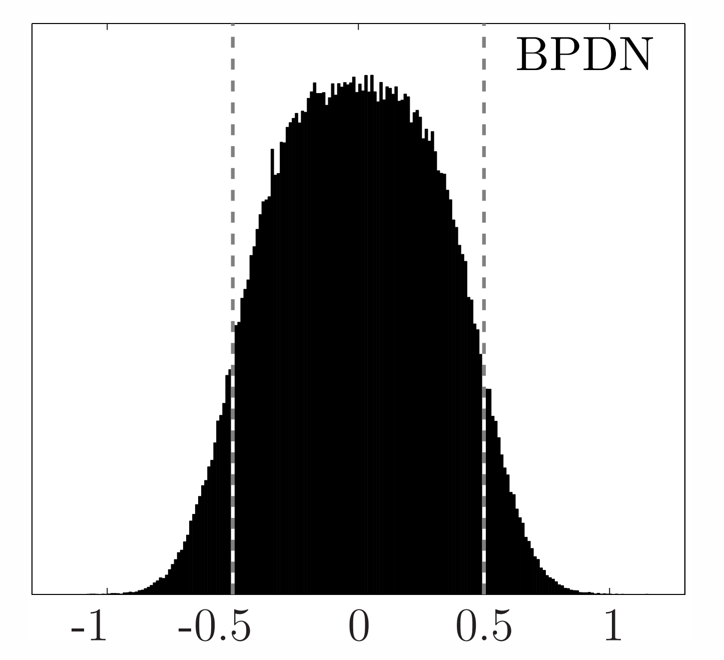
Dequantizing CS?

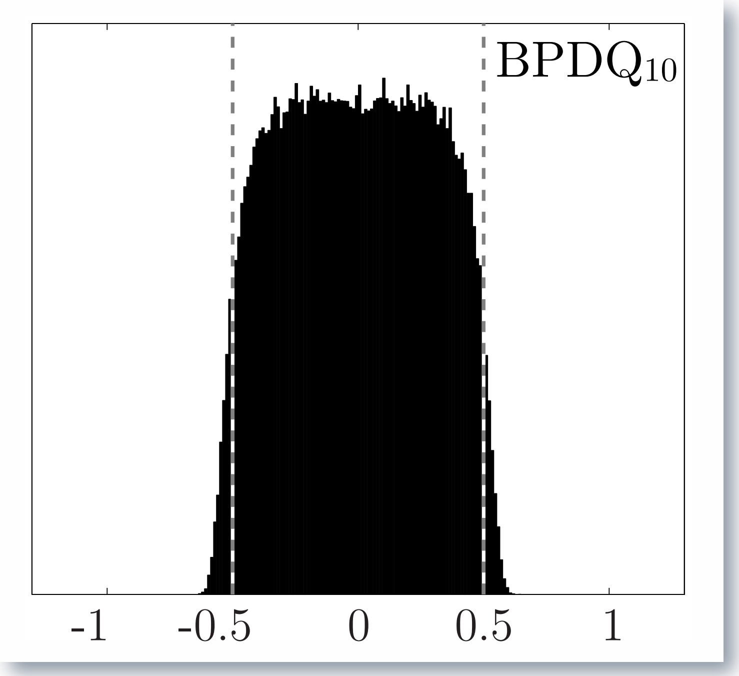
[LJ, Hammond, Fadili, 2009, 2011]



* N=1024, K=16, Gaussian $\boldsymbol{\Phi}$
* 500 K-sparse (canonical basis)
* Non-zero components follow $\mathcal{N}(0,1)$
* Quantiz. bin width $\alpha = \|\boldsymbol{\Phi x}\|_\infty/40$

Histograms of
$\alpha^{-1}(\boldsymbol{q} - \boldsymbol{\Phi}\hat{\boldsymbol{x}})_i$



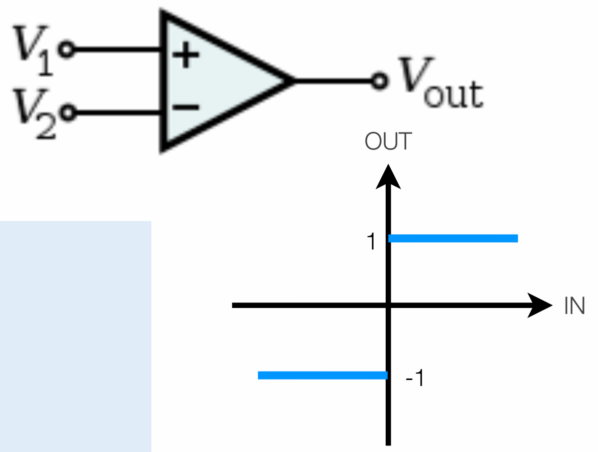


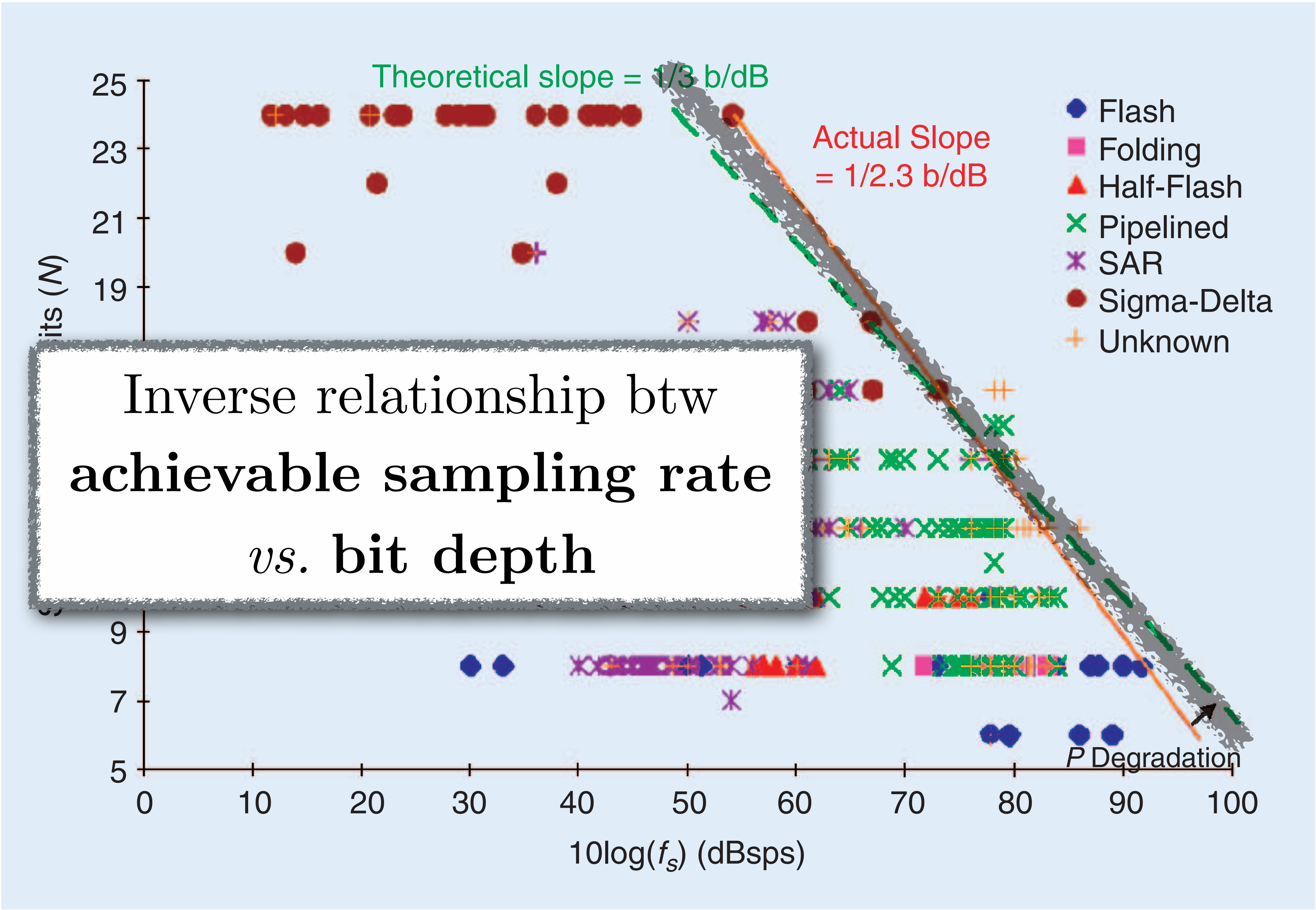
LJ, D. Hammond, J. Fadili "Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine." Information Theory, IEEE Transactions on, 57(1), 559-571.

V. 1-bit CS and Binary ε-Stable Embedding

(first observed quasi-isometry in CS)

Why 1-bit? Very Fast Quantizers!



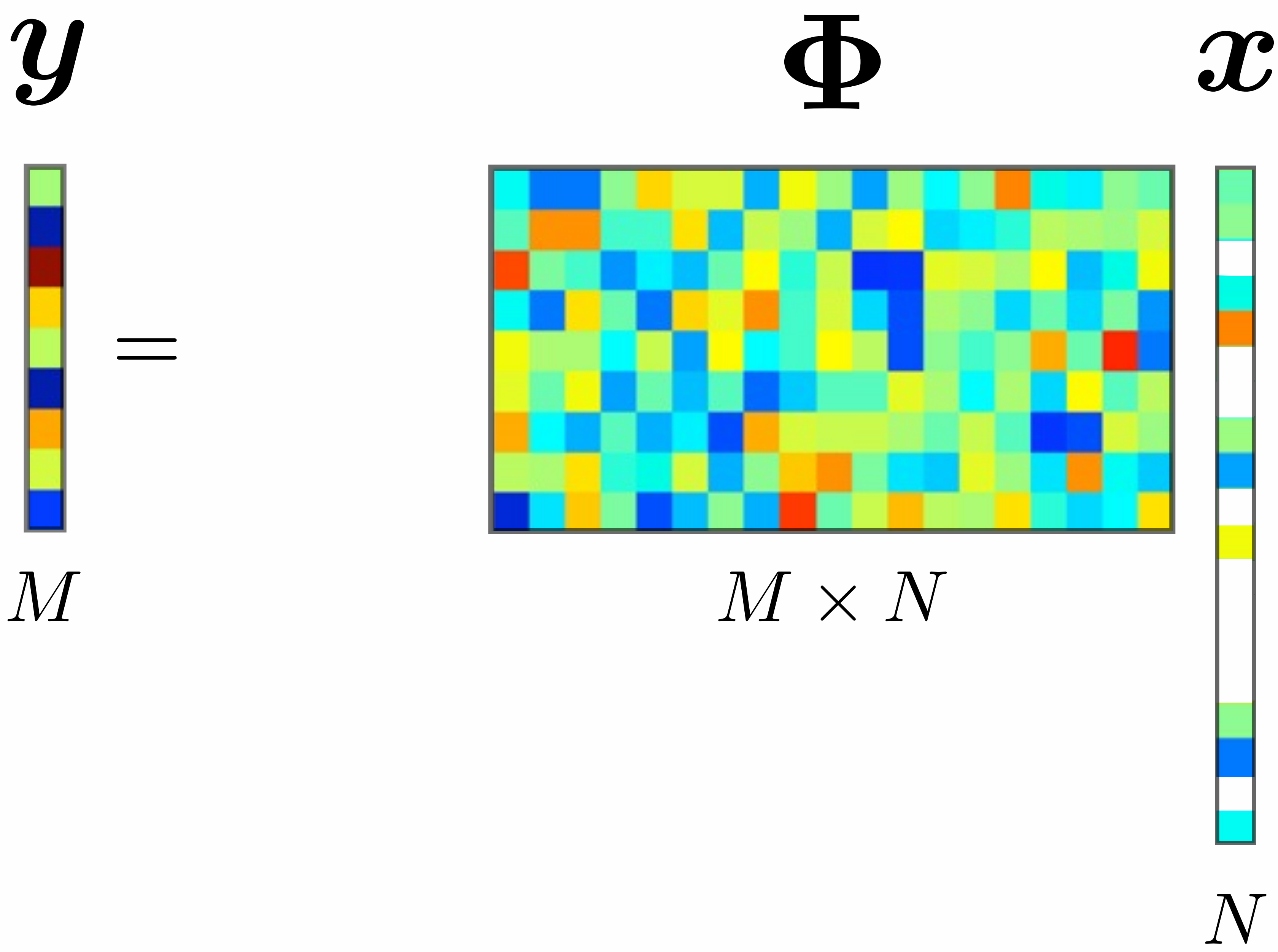


Inverse relationship btw **achievable sampling rate** *vs.* **bit depth**

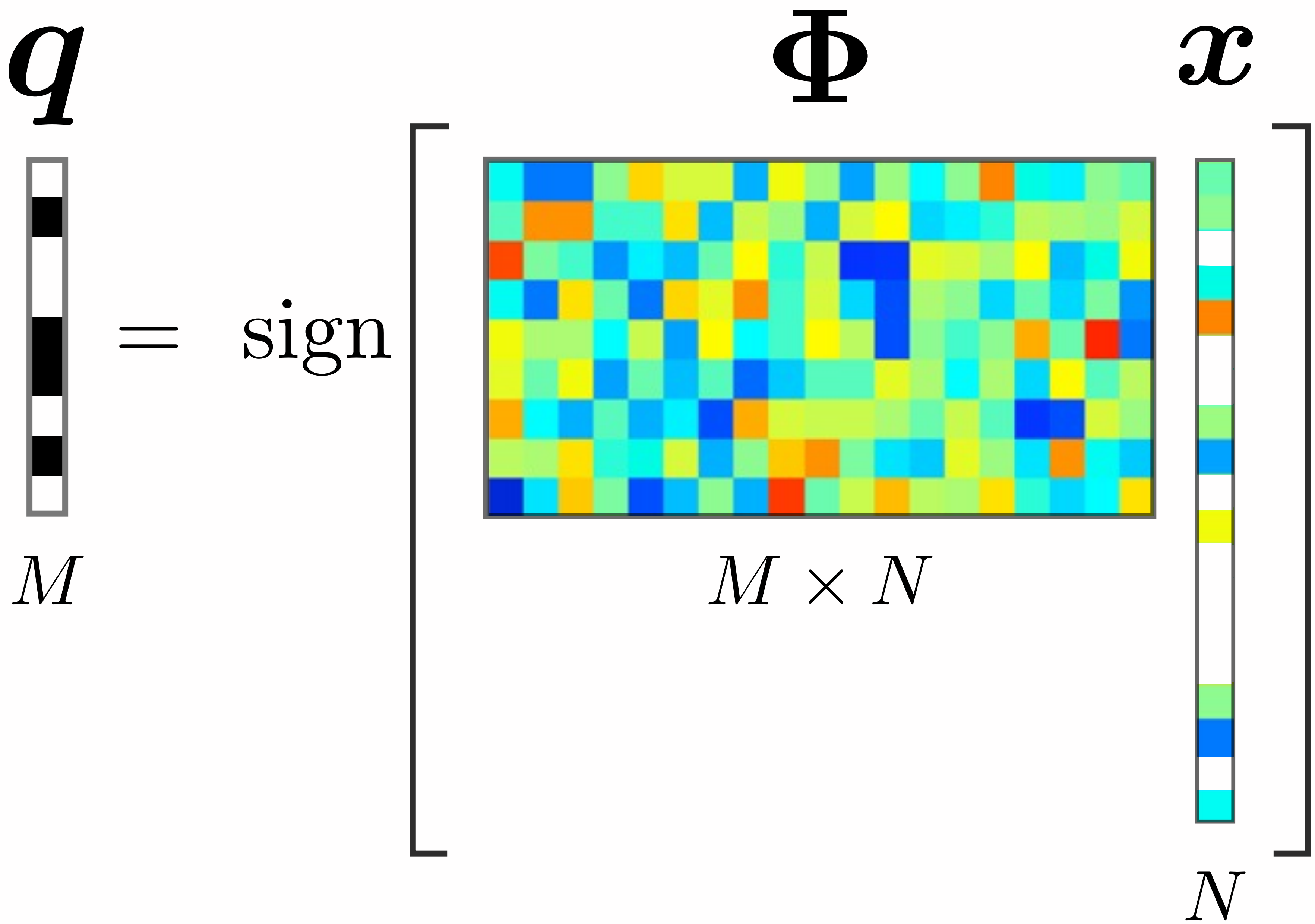
[FIG1] Stated number of bits versus sampling rate.

[From "Analog-to-digital converters" B. Le, T.W. Rondeau, J.H. Reed, and C.W.Bostian, IEEE Sig. Proc. Magazine, Nov 2005]

1-bit Compressed Sensing



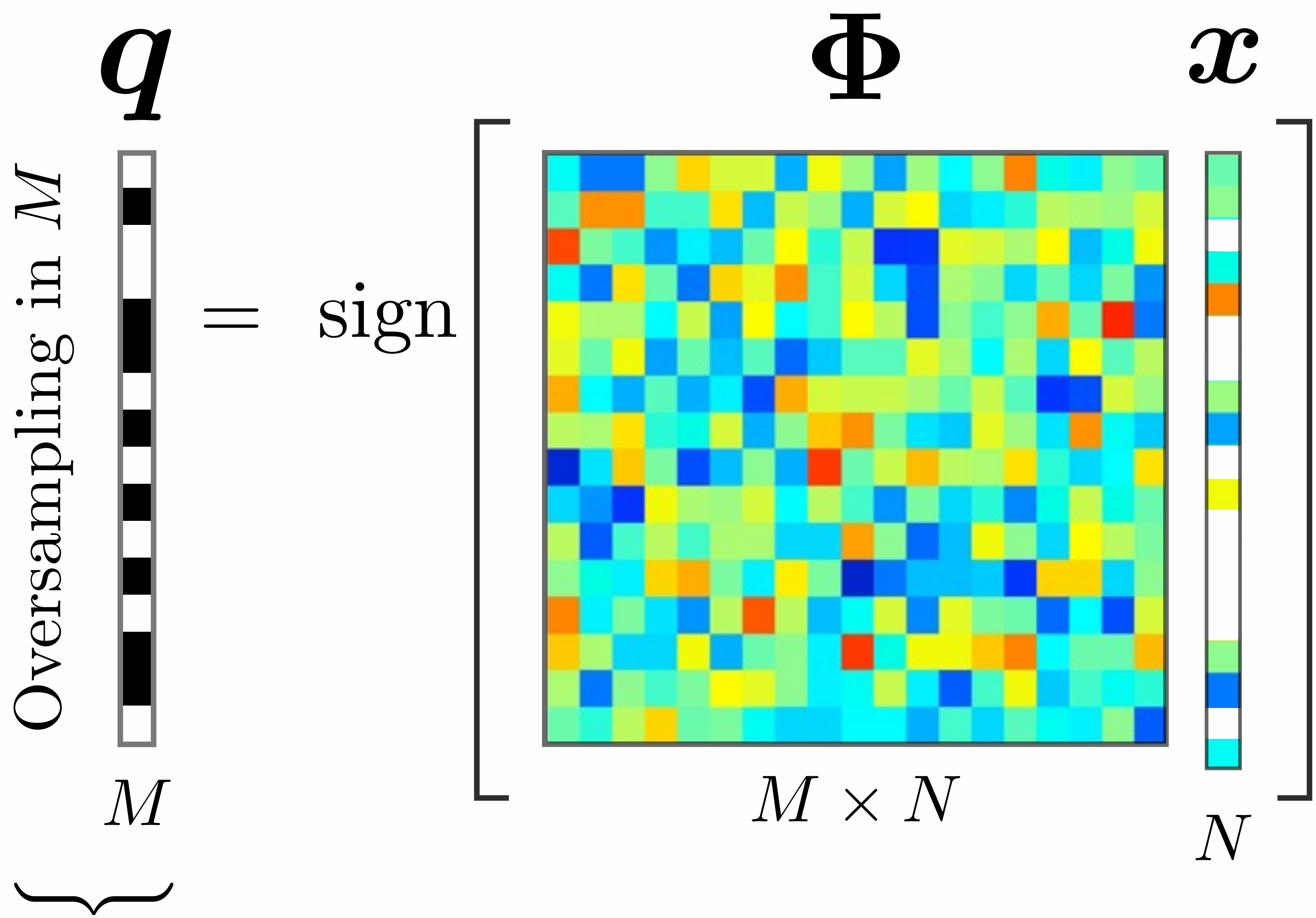
1-bit Compressed Sensing



with: $\operatorname{sign} t = \begin{cases} 1 & \text{if } t > 0 \\ -1 & \text{if } t \leqslant 0 \end{cases}$ component-wise

1-bit ~~Compressed~~ Computational Sensing

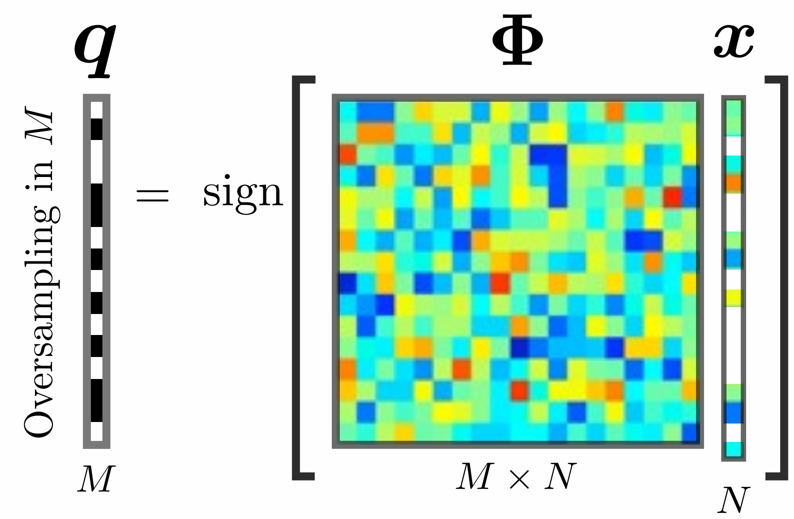
bits matter!



M-bits! But, which information inside $\boldsymbol{q}$?

1-bit ~~Compressed~~ Computational Sensing

bits matter!



<u>Warning 1</u>: signal amplitude is lost!

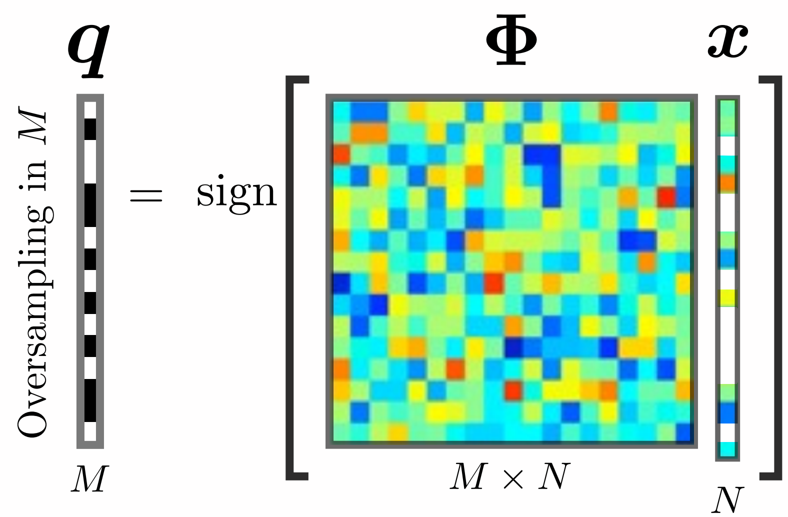
$$\boldsymbol{q} = \operatorname{sign}\left(\boldsymbol{\Phi}(\lambda \boldsymbol{x})\right) = \operatorname{sign}\left(\boldsymbol{\Phi}\boldsymbol{x}\right), \quad \forall \lambda > 0$$

$\Rightarrow$ Amplitude is arbitrarily fixed

Examples : $\|\boldsymbol{x}\| = 1$ or $\|\boldsymbol{\Phi}\boldsymbol{x}\|_1 = 1$

1-bit ~~Compressed~~ Computational Sensing

bits matter!



[Plan, Vershynin, 11]

Warning 2: $\exists$ forbidden sensing!

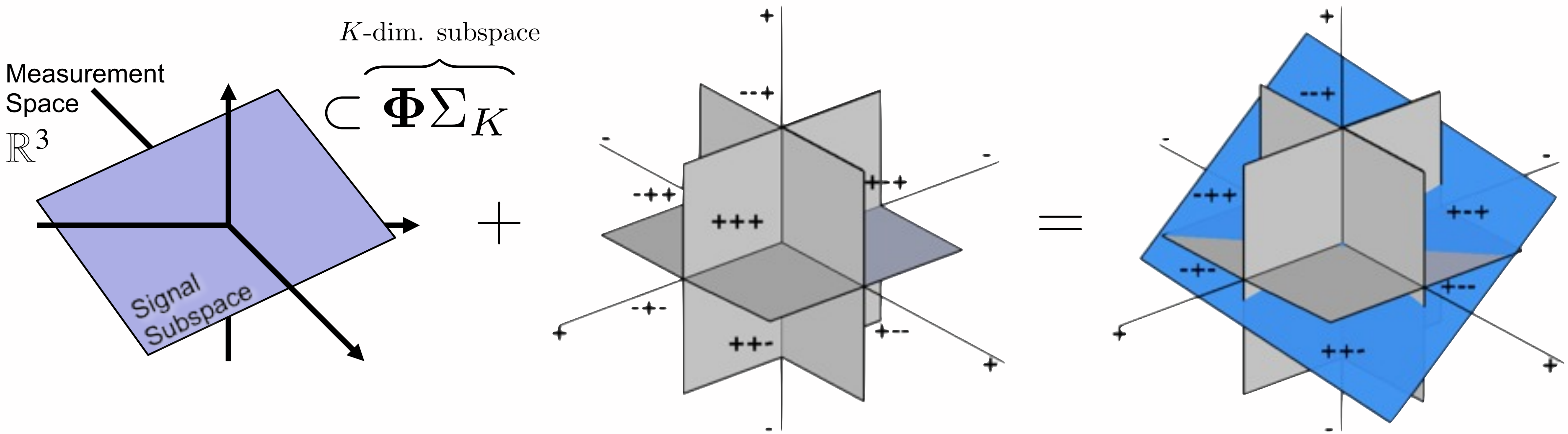
Let $\boldsymbol{x}_\lambda := (1, \lambda, 0, \cdots, 0)^T \in \mathbb{R}^N$
and $\boldsymbol{\Phi} \in \{\pm 1\}^{M \times N}$ (*e.g.*, Bernoulli).

We have $\|\boldsymbol{x}_0 - \boldsymbol{x}_\lambda\| = \lambda$
but $\boldsymbol{q} = \mathrm{sign}\,(\boldsymbol{\Phi}\boldsymbol{x}_0) = \mathrm{sign}\,(\boldsymbol{\Phi}\boldsymbol{x}_\lambda),\ \forall |\lambda| < 1$

$\Rightarrow$ No hope to distinguish them by increasing M!

beware of "too sparse" vector difference!! (see later)

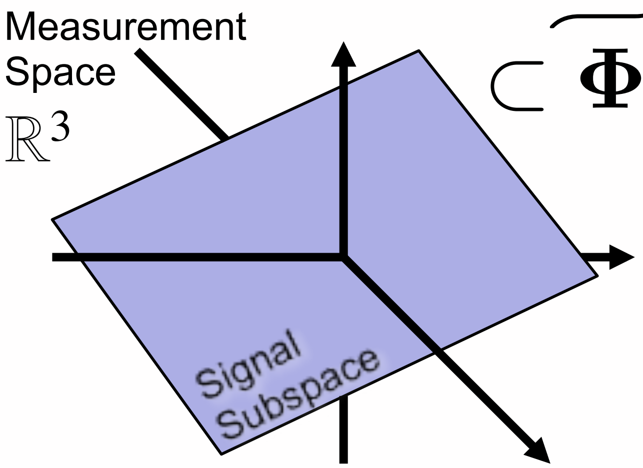
Lower Bound on Reconstruction Error



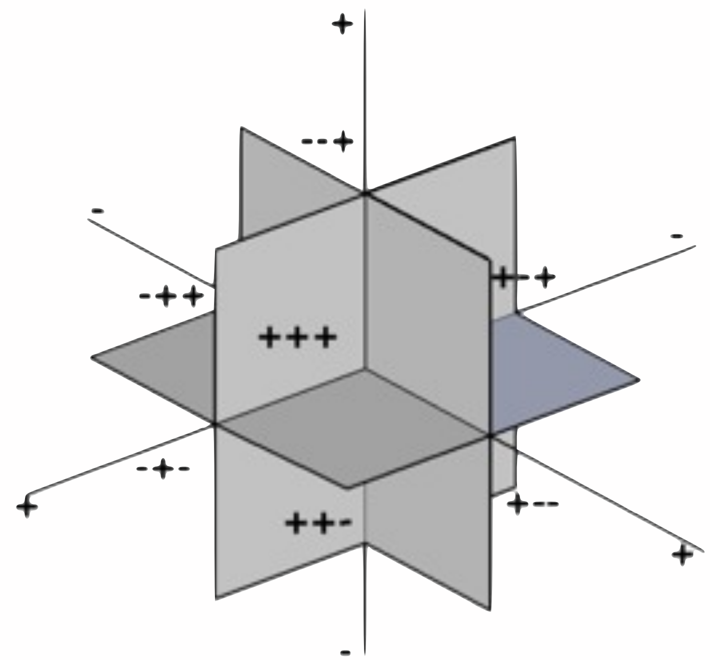
Not all quantization cells intersected!

no more than $C = 2^K \binom{N}{K}\binom{M}{K}$ (e.g., [Cover 65, Flatto 70])

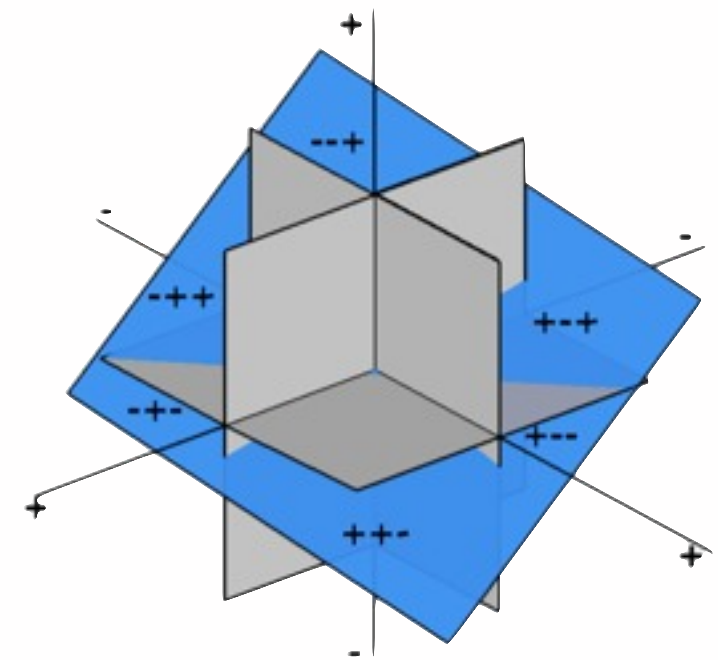
Lower Bound on Reconstruction Error



$\subset \overbrace{\boldsymbol{\Phi}\Sigma_K}^{K\text{-dim. subspace}}$ $+$



$=$

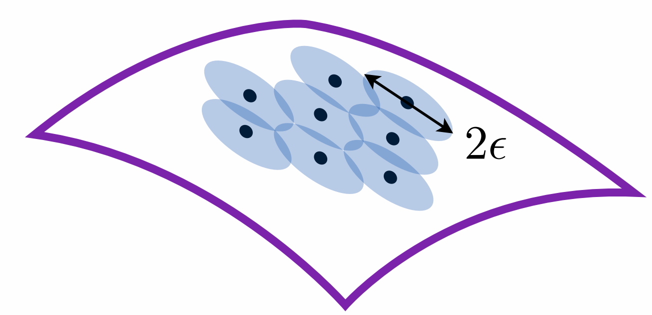


Not all quantization cells intersected!

no more than $C = 2^K \binom{N}{K}\binom{M}{K}$ (e.g., [Cover 65, Flatto 70])

For an error $\epsilon > 0$,

$$\boxed{C \,\mathrm{vol}(\epsilon\text{−cap}) \geqslant \mathrm{vol}(S^{N-1} \cap \Sigma_K)}$$



and $\mathrm{vol}(S^{N-1} \cap \Sigma_K)/\mathrm{vol}(\epsilon\text{−cap}) \simeq \binom{N}{K}\epsilon^{-K}$

Lower bound on any 1-bit reconstruction error: $\Rightarrow \epsilon = \Omega(K/M)$

Reaching this bound?

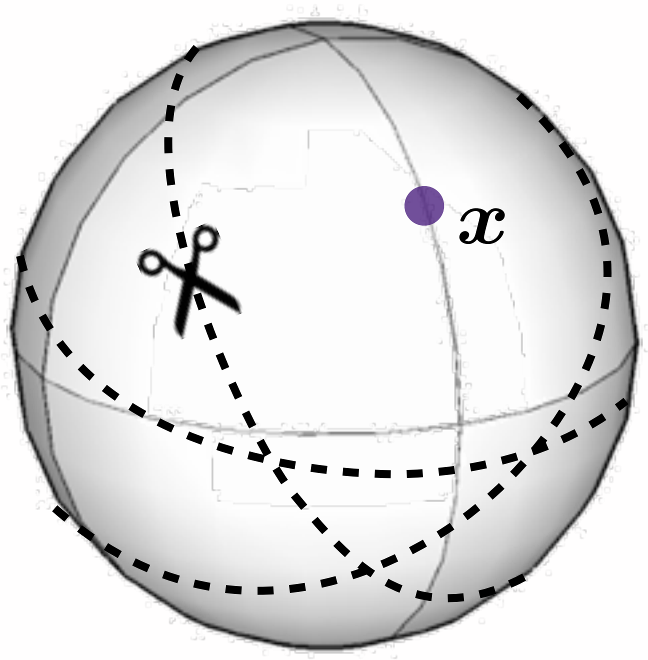
Reaching this bound?

$\boldsymbol{x}$ on S^2

M vectors:

$\{\boldsymbol{\varphi}_i : 1 \leqslant i \leqslant M\}$

iid Gaussian



[Illustration: P. Boufounos]

Reaching this bound?

$\boldsymbol{x}$ on S^2

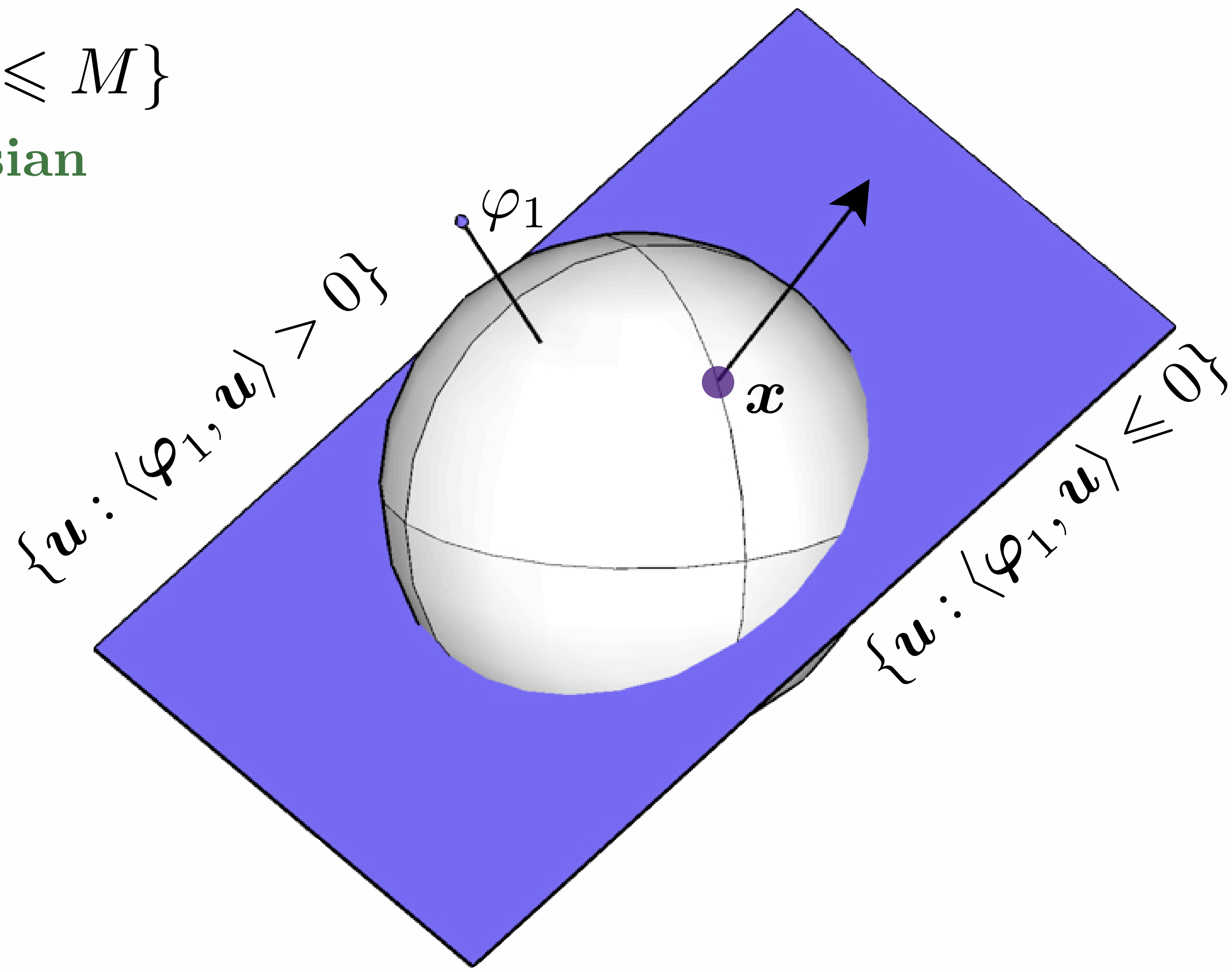
M vectors:

$\{\boldsymbol{\varphi}_i : 1 \leqslant i \leqslant M\}$

iid Gaussian

1-bit Measurements

$\langle \boldsymbol{\varphi}_1, \boldsymbol{x} \rangle > 0$



[Illustration: P. Boufounos]

Reaching this bound?

$\boldsymbol{x}$ on S^2

M vectors:

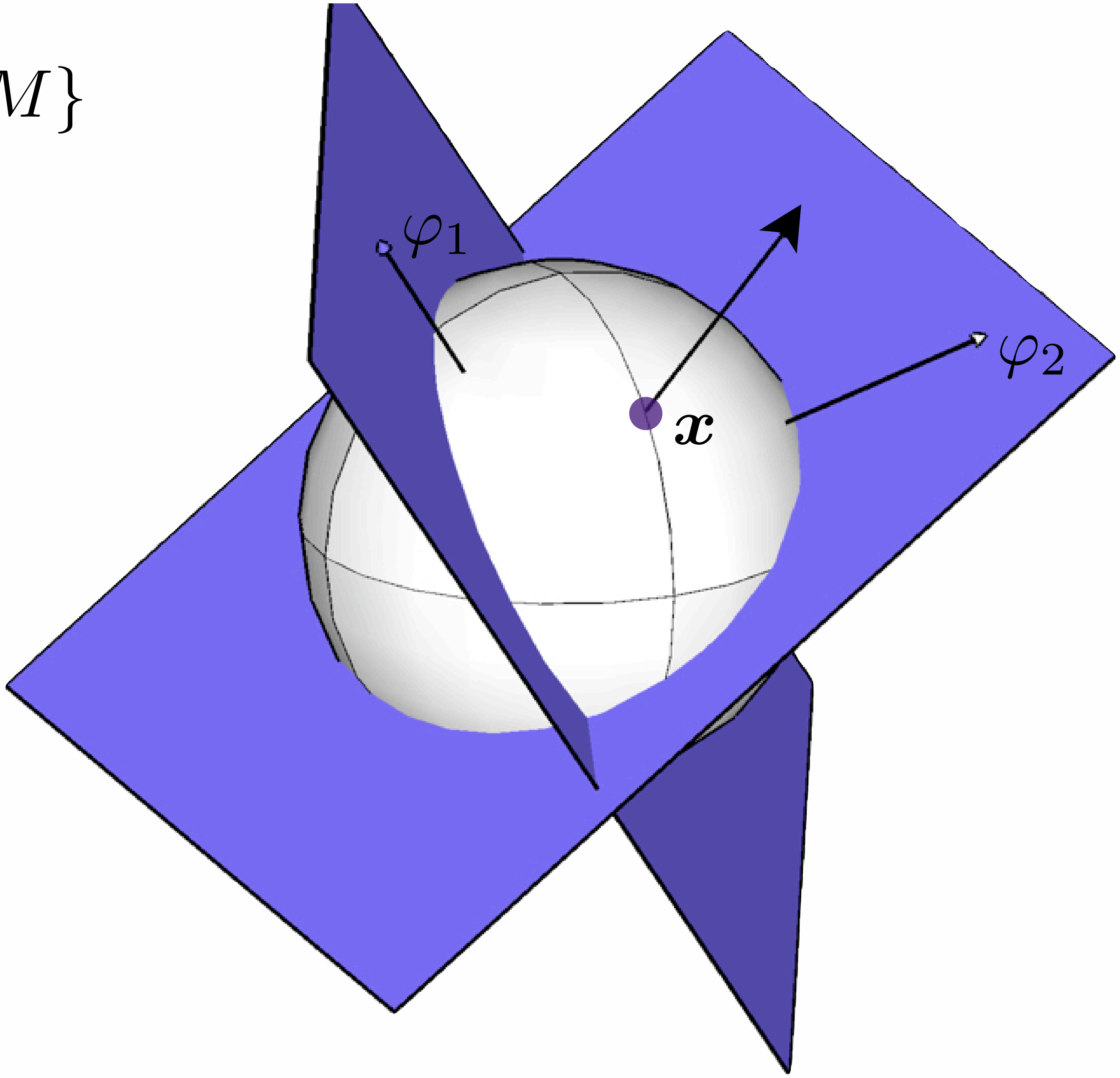
$\{\boldsymbol{\varphi}_i : 1 \leqslant i \leqslant M\}$

iid Gaussian

1-bit Measurements

$\langle \boldsymbol{\varphi}_1, \boldsymbol{x} \rangle > 0$

$\langle \boldsymbol{\varphi}_2, \boldsymbol{x} \rangle > 0$



[Illustration: P. Boufounos]

Reaching this bound?

$\boldsymbol{x}$ on S^2

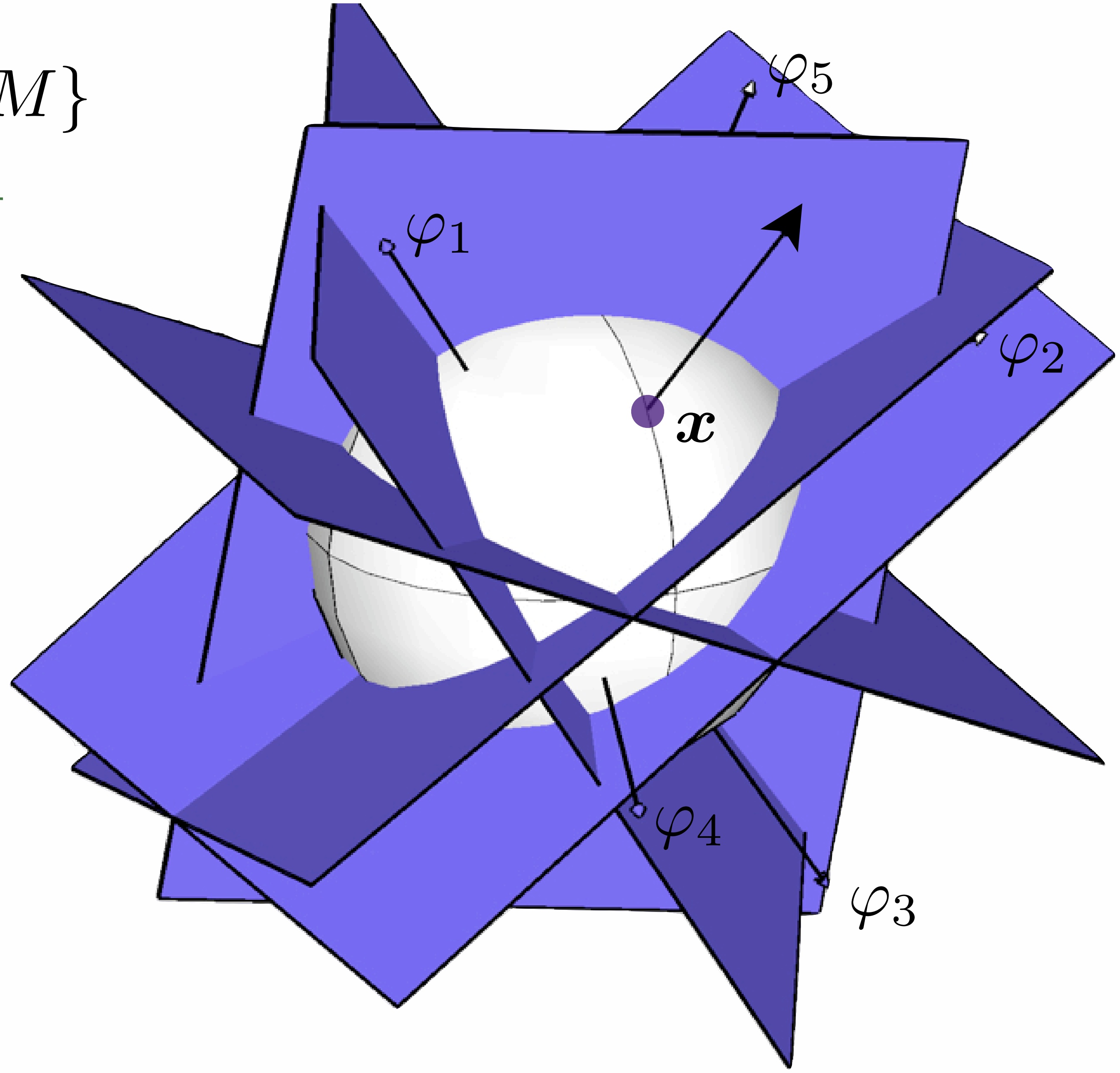
M vectors:

$\{\boldsymbol{\varphi}_i : 1 \leqslant i \leqslant M\}$

iid Gaussian

1-bit Measurements

$\langle \boldsymbol{\varphi}_1, \boldsymbol{x} \rangle > 0$
$\langle \boldsymbol{\varphi}_2, \boldsymbol{x} \rangle > 0$
$\langle \boldsymbol{\varphi}_3, \boldsymbol{x} \rangle \leqslant 0$
$\langle \boldsymbol{\varphi}_4, \boldsymbol{x} \rangle > 0$
$\langle \boldsymbol{\varphi}_5, \boldsymbol{x} \rangle > 0$



[Illustration: P. Boufounos]

Reaching this bound?

$\boldsymbol{x}$ on S^2

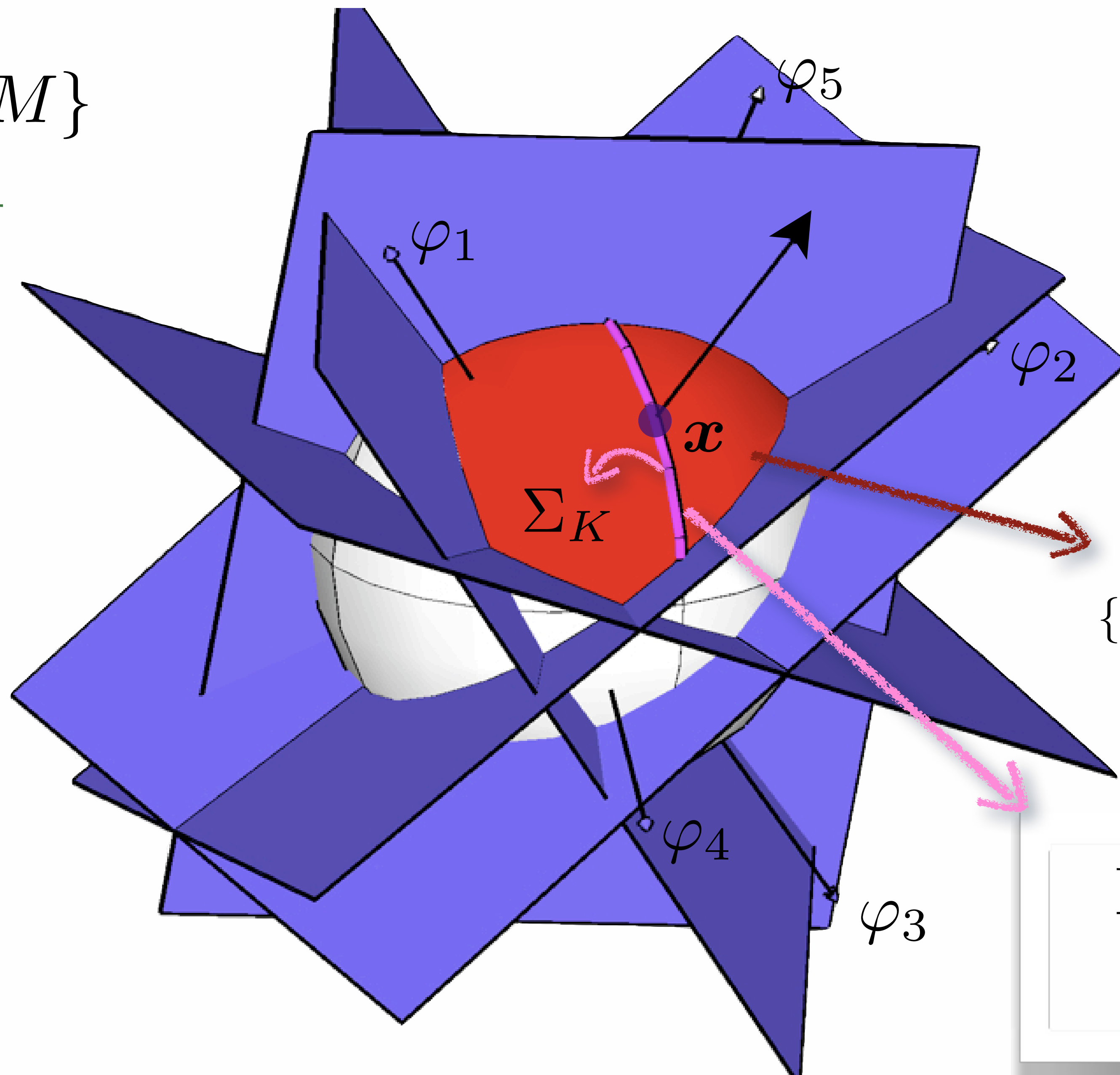
M vectors:

$\{\boldsymbol{\varphi}_i : 1 \leqslant i \leqslant M\}$

iid Gaussian

1-bit Measurements

$\langle \boldsymbol{\varphi}_1, \boldsymbol{x} \rangle > 0$
$\langle \boldsymbol{\varphi}_2, \boldsymbol{x} \rangle > 0$
$\langle \boldsymbol{\varphi}_3, \boldsymbol{x} \rangle \leqslant 0$
$\langle \boldsymbol{\varphi}_4, \boldsymbol{x} \rangle > 0$
$\langle \boldsymbol{\varphi}_5, \boldsymbol{x} \rangle > 0$
$\vdots$



Smaller and smaller when M increases

$\{\boldsymbol{u} : \text{sign}\,(\boldsymbol{\Phi u}) = \text{sign}\,(\boldsymbol{\Phi x})\}$

Lower bound on this width?

[Illustration: P. Boufounos]

Reaching this bound?

Let $A(\cdot) := \mathrm{sign}\,(\boldsymbol{\Phi}\,\cdot)$ with $\boldsymbol{\Phi} \sim \mathcal{N}^{M\times N}(0,1)$.

If $M = O(\epsilon^{-1} K \log N\,)$, then, w.h.p,

for any two unit K-sparse vectors $\boldsymbol{x}$ and $\boldsymbol{s}$,

$$A(\boldsymbol{x}) = A(\boldsymbol{s}) \quad \Rightarrow \quad \|\boldsymbol{x} - \boldsymbol{s}\| \leq \epsilon$$

$$\Leftrightarrow \epsilon = O\big(\tfrac{K}{M} \log \tfrac{MN}{K}\big)$$

almost optimal

Reaching this bound?

Let $A(\cdot) := \mathrm{sign}\,(\boldsymbol{\Phi}\,\cdot)$ with $\boldsymbol{\Phi} \sim \mathcal{N}^{M\times N}(0,1)$.

If $M = O(\epsilon^{-1} K \log N)$, then, w.h.p,

for any two unit K-sparse vectors $\boldsymbol{x}$ and $\boldsymbol{s}$,

$$A(\boldsymbol{x}) = A(\boldsymbol{s}) \quad \Rightarrow \quad \|\boldsymbol{x} - \boldsymbol{s}\| \leq \epsilon$$

$$\Leftrightarrow \epsilon = O\big(\tfrac{K}{M} \log \tfrac{MN}{K}\big)$$

almost optimal

<u>Note</u>: You can even afford a small error [LJ, Degraux 2013], *i.e.*,

if only b bits are different between $A(\boldsymbol{x})$ and $A(\boldsymbol{s})$ $\Rightarrow \|\boldsymbol{x} - \boldsymbol{s}\| \leqslant \frac{K+b}{K}\,\epsilon$

Embeddings?

Central question:

(Hamming) distance between $A(\boldsymbol{x})$ and $A(\boldsymbol{s})$

$$\simeq$$

(angular) distance between $\boldsymbol{x}$ and $\boldsymbol{s}$?

Uniformly for all K-sparse vectors?

Distances of interest:

$$d_H(\boldsymbol{u}, \boldsymbol{v}) = \tfrac{1}{M} \textstyle\sum_i (u_i \oplus v_i) \quad \text{(norm. Hamming)}$$

$$d_{\text{ang}}(\boldsymbol{x}, \boldsymbol{s}) = \tfrac{1}{\pi} \arccos(\langle \boldsymbol{x}, \boldsymbol{s} \rangle) \quad \text{(norm. angle)}$$

Binary ϵ - Stable Embedding (BϵSE)

Definition:

A mapping $A : \mathbb{R}^N \to \{\pm 1\}^M$ is a **binary ϵ-stable embedding** (BϵSE) of order K for sparse vectors if

$$d_{\text{ang}}(\boldsymbol{x}, \boldsymbol{s}) - \epsilon \leq d_H(A(\boldsymbol{x}), A(\boldsymbol{s})) \leq d_{\text{ang}}(\boldsymbol{x}, \boldsymbol{s}) + \epsilon$$

for all $\boldsymbol{x}, \boldsymbol{s} \in S^{N-1}$ with $\boldsymbol{x} \pm \boldsymbol{s}$ K-sparse.

kind of "binary restricted (quasi) isometry"

Binary ϵ - Stable Embedding (BϵSE)

Definition:

A mapping $A : \mathbb{R}^N \to \{\pm 1\}^M$ is a **binary ϵ-stable embedding** (BϵSE) of order K for sparse vectors if

$$d_{\rm ang}(\boldsymbol{x}, \boldsymbol{s}) - \epsilon \leq d_H(A(\boldsymbol{x}), A(\boldsymbol{s})) \leq d_{\rm ang}(\boldsymbol{x}, \boldsymbol{s}) + \epsilon$$

for all $\boldsymbol{x}, \boldsymbol{s} \in S^{N-1}$ with $\boldsymbol{x} \pm \boldsymbol{s}$ K-sparse.

kind of "binary restricted (quasi) isometry"

Corollary: for any algorithm with output $\boldsymbol{x}^*$,

jointly K-sparse and consistent (*i.e.*, $A(\boldsymbol{x}^*) = A(\boldsymbol{x})$), we have:

$$d_{\rm ang}(\boldsymbol{x}, \boldsymbol{x}^*) \leqslant 2\epsilon!$$

- If limited binary noise, $d_{\rm ang}$ still bounded
- If not exactly sparse signals (but almost), $d_{\rm ang}$ still bounded

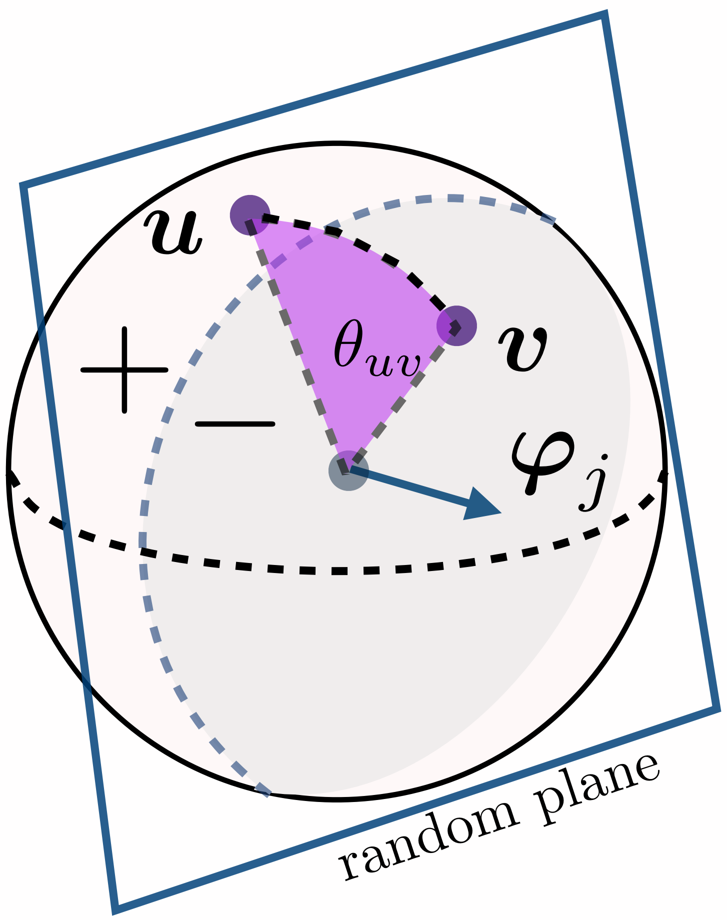
BϵSE existence? Yes!

Let $\boldsymbol{\Phi} \sim \mathcal{N}^{M\times N}(0,1)$, fix $0 \leq \eta \leq 1$ and $\epsilon > 0$. If $\boldsymbol{A} := \operatorname{sign}(\boldsymbol{\Phi}\cdot)$ and

$$M \gtrsim \tfrac{1}{\epsilon^2} K \log \tfrac{N}{\epsilon} + \log \tfrac{1}{\eta},$$

then $\boldsymbol{A}$ is a BϵSE with Pr $> 1 - \eta$.

$$M = O(\epsilon^{-2} K \log N)$$



Proof sketch:

[*e.g.*, Goemans, Williamson 1995]

Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{S}^{N-1}$ (wlog) and $A_j(\cdot) = \operatorname{sign}(\boldsymbol{\varphi}_j^T \cdot)$

$$\mathbb{P}[A_j(\boldsymbol{u}) \neq A_j(\boldsymbol{v})] = \tfrac{1}{\pi}\,\mathrm{angle}(\boldsymbol{u}, \boldsymbol{v}) = \tfrac{1}{\pi}\theta_{uv}$$

+ measure concentration on $d_H \sim \mathrm{Bin}(\frac{1}{\pi}\theta_{uv}, M)$
+ covering/approximate continuity

VI. Quantized Embeddings

1. Beyond strict sparsity ... [Plan, Vershynin]

Let $\mathcal{K} \subset S^{N-1}$ (*e.g.*, compressible signals s.t. $\|\boldsymbol{x}\|_2/\|\boldsymbol{x}\|_1 \leqslant \sqrt{K}$)
$\neq \Sigma_K$

What can we say on $d_H(A(\boldsymbol{x}), A(\boldsymbol{s}))$ for $\boldsymbol{x}, \boldsymbol{s} \in \mathcal{K}$?

Y. Plan, R. Vershynin, "Dimension reduction by random hyperplane tessellations", 2011, arXiv:1111.4452
Y. Plan, R. Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: a convex programming approach", IEEE TIT 2012, arXiv:1202.1212.

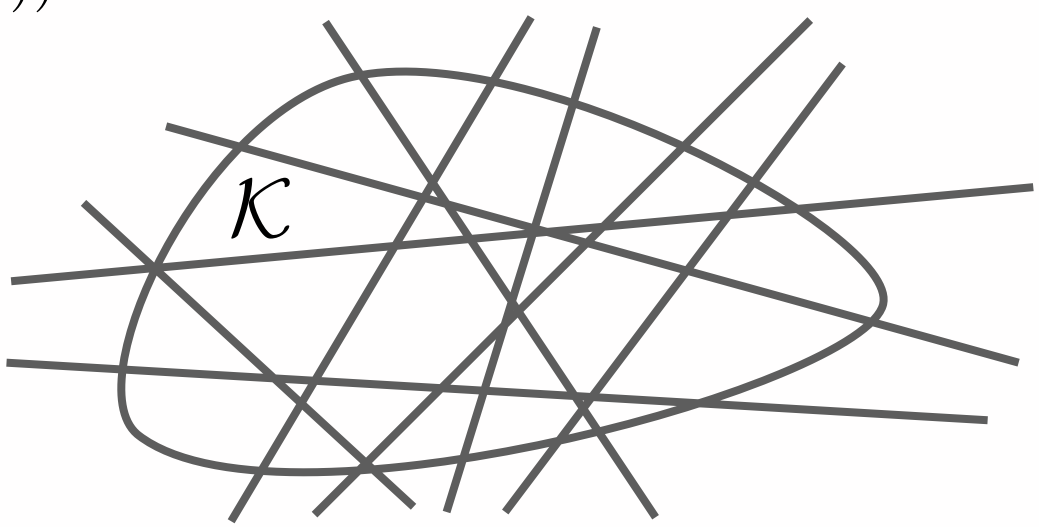
1. Beyond strict sparsity ... [Plan, Vershynin]

Let $\mathcal{K} \subset S^{N-1}$ (*e.g.*, compressible signals s.t. $\|\boldsymbol{x}\|_2/\|\boldsymbol{x}\|_1 \leqslant \sqrt{K}$)
$\neq \Sigma_K$

What can we say on $d_H(A(\boldsymbol{x}), A(\boldsymbol{s}))$ for $\boldsymbol{x}, \boldsymbol{s} \in \mathcal{K}$?

Uniform tessellation: [Plan, Vershynin, 11]

$$\mathrm{P}\big(\underbrace{\#\text{ random hyperplanes btw } \boldsymbol{x}}_{d_H(A(\boldsymbol{x}), A(\boldsymbol{s}))} \text{ and } \boldsymbol{s} \propto d_{\rm ang}(\boldsymbol{x}, \boldsymbol{s})\big)\ ?$$

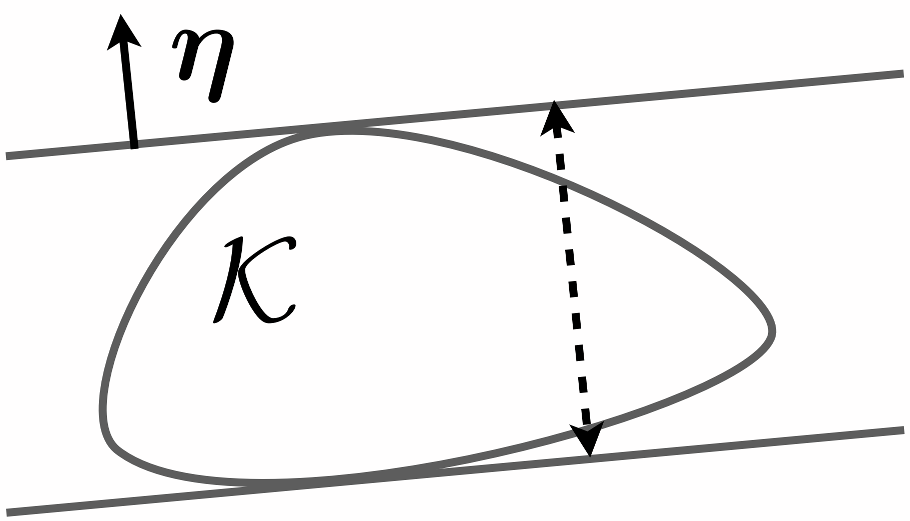


Y. Plan, R. Vershynin, "Dimension reduction by random hyperplane tessellations", 2011, arXiv:1111.4452
Y. Plan, R. Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: a convex programming approach", IEEE TIT 2012, arXiv:1202.1212.

1. Beyond strict sparsity ... [Plan, Vershynin]

Measuring the "dimension" of $\mathcal{K}$ $\to$ Gaussian mean width:

$$w(\mathcal{K}) := \mathbb{E} \sup_{\boldsymbol{u} \in \mathcal{K}} |\langle \boldsymbol{g}, \boldsymbol{u} \rangle|, \text{ with } g_k \sim_{\text{iid}} \mathcal{N}(0,1)$$



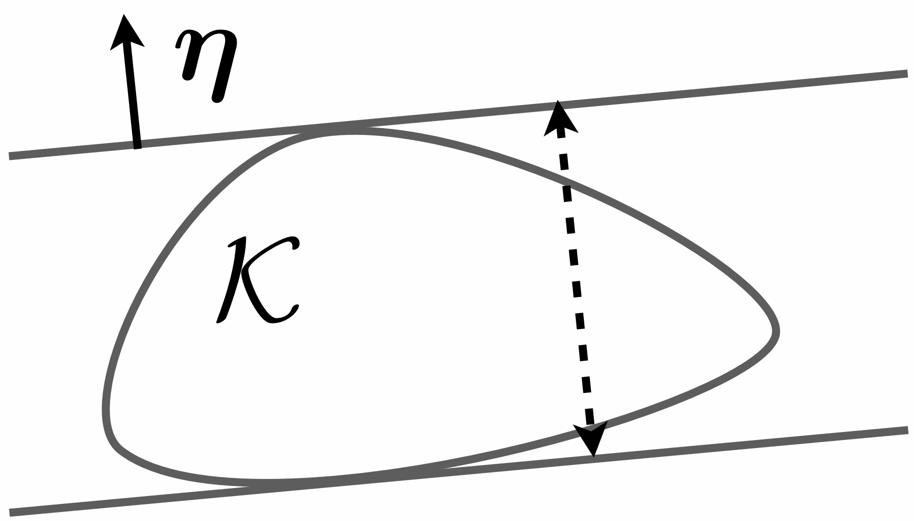
width in direction $\boldsymbol{\eta}$

Y. Plan, R. Vershynin, "Dimension reduction by random hyperplane tessellations", 2011, arXiv:1111.4452
Y. Plan, R. Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: a convex programming approach", IEEE TIT 2012, arXiv:1202.1212.

1. Beyond strict sparsity ... [Plan, Vershynin]

Measuring the "dimension" of $\mathcal{K} \to$ Gaussian mean width:

$$w(\mathcal{K}) := \mathbb{E} \sup_{\boldsymbol{u} \in \mathcal{K}} |\langle \boldsymbol{g}, \boldsymbol{u} \rangle|, \text{ with } g_k \sim_{\text{iid}} \mathcal{N}(0,1)$$



width in direction $\boldsymbol{\eta}$

Examples:

$w^2(\mathcal{S}^{N-1}) \leqslant 4N$

$w^2(\mathcal{K}) \leqslant C \log |\mathcal{K}|$ (for finite sets)

$w^2(\mathcal{K}) \leqslant L$ if subspace with $\dim \mathcal{K} = L$

$w^2(\Sigma_K) \simeq K \log(2N/K)$

$\vdots$

Y. Plan, R. Vershynin, "Dimension reduction by random hyperplane tessellations", 2011, arXiv:1111.4452

Y. Plan, R. Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: a convex programming approach", IEEE TIT 2012, arXiv:1202.1212.

1. Beyond strict sparsity ... [Plan, Vershynin]

Proposition *Let* $\mathbf{\Phi} \sim \mathcal{N}^{M\times N}(0,1)$ *and* $\mathcal{K} \subset \mathbb{R}^N$. *Then, for some* $C, c > 0$, *if*

$$M \ \geqslant\ C\epsilon^{-6} w^2(\mathcal{K}),$$

then, with $Pr \geqslant 1 - e^{-c\epsilon^2 M}$, *we have*

$$d_{\rm ang}(\boldsymbol{x},\boldsymbol{s}) - \epsilon \ \leqslant\ d_H(A(\boldsymbol{x}),A(\boldsymbol{s})) \ \leqslant\ d_{\rm ang}(\boldsymbol{x},\boldsymbol{s}) - \epsilon, \quad \forall \boldsymbol{x},\boldsymbol{s} \in \mathcal{K}.$$

Y. Plan, R. Vershynin, "Dimension reduction by random hyperplane tessellations", 2011, arXiv:1111.4452
Y. Plan, R. Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: a convex programming approach", IEEE TIT 2012, arXiv:1202.1212.

1. Beyond strict sparsity ... [Plan, Vershynin]

Proposition *Let* $\boldsymbol{\Phi} \sim \mathcal{N}^{M\times N}(0,1)$ *and* $\mathcal{K} \subset \mathbb{R}^N$. *Then, for some* $C, c > 0$, *if*

$$M \geqslant C\epsilon^{-6} w^2(\mathcal{K}),$$

not as optimal but stronger result!

then, with $Pr \geqslant 1 - e^{-c\epsilon^2 M}$, *we have*

$$d_{\rm ang}(\boldsymbol{x}, \boldsymbol{s}) - \epsilon \leqslant d_H(A(\boldsymbol{x}), A(\boldsymbol{s})) \leqslant d_{\rm ang}(\boldsymbol{x}, \boldsymbol{s}) - \epsilon, \quad \forall \boldsymbol{x}, \boldsymbol{s} \in \mathcal{K}.$$

BϵSE is generalized to more general sets.

In particular, to

$$\mathcal{C}_K = \{\boldsymbol{u} \in \mathbb{R}^N : \|\boldsymbol{u}\|_1 \leq \sqrt{K}, \|\boldsymbol{u}\| \leq 1\} \supset \Sigma_K \cap \mathbb{B}^N$$

with $w^2(\mathcal{C}_K) \leq cK \log N/K$.

Y. Plan, R. Vershynin, "Dimension reduction by random hyperplane tessellations", 2011, arXiv:1111.4452

Y. Plan, R. Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: a convex programming approach", IEEE TIT 2012, arXiv:1202.1212.

1. Beyond strict sparsity ... [Plan, Vershynin]

Proposition *Let* $\boldsymbol{\Phi} \sim \mathcal{N}^{M\times N}(0,1)$ *and* $\mathcal{K} \subset \mathbb{R}^N$*. Then, for some* $C, c > 0$*, if*

$$M \geqslant C\epsilon^{-6} w^2(\mathcal{K}),$$

not as optimal but stronger result!

then, with $Pr \geqslant 1 - e^{-c\epsilon^2 M}$*, we have*

$$d_{\rm ang}(\boldsymbol{x}, \boldsymbol{s}) - \epsilon \leqslant d_H(A(\boldsymbol{x}), A(\boldsymbol{s})) \leqslant d_{\rm ang}(\boldsymbol{x}, \boldsymbol{s}) - \epsilon, \quad \forall \boldsymbol{x}, \boldsymbol{s} \in \mathcal{K}.$$

BϵSE is generalized to more general sets.

In particular, to

$$\mathcal{C}_K = \{\boldsymbol{u} \in \mathbb{R}^N : \|\boldsymbol{u}\|_1 \leq \sqrt{K}, \|\boldsymbol{u}\| \leq 1\} \supset \Sigma_K \cap \mathbb{B}^N$$

with $w^2(\mathcal{C}_K) \leq cK \log N/K$.

$\Rightarrow$ Extension to "1-bit Matrix Completion" possible!

i.e., $w^2(r\text{-rank } N_1 \times N_2 \text{ matrix}) \leqslant c\, r(N_1 + N_2)$!

Y. Plan, R. Vershynin, "Dimension reduction by random hyperplane tessellations", 2011, arXiv:1111.4452

Y. Plan, R. Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: a convex programming approach", IEEE TIT 2012, arXiv:1202.1212.

2. Beyond 1-bit ... Quantizing JL lemma?

- The Johnson-Lindenstrauss Lemma (1984)

2. Beyond 1-bit … Quantizing JL lemma?

- The Johnson-Lindenstrauss Lemma (1984)

Lemma 1 *Given an error* $0 < \epsilon < 1$, *and a point set* $\mathcal{S} \subset \mathbb{R}^N$. *If* M *is such that*

$$M > M_0 = O(\epsilon^{-2} \log |\mathcal{S}|),$$

then, there exists a (Lipschitz) mapping $\boldsymbol{f} : \mathbb{R}^N \to \mathbb{R}^M$ *such that*

$$(1-\epsilon)\,\|\boldsymbol{u}-\boldsymbol{v}\| \;\leqslant\; \tfrac{1}{\sqrt{M}}\,\|\boldsymbol{f}(\boldsymbol{u})-\boldsymbol{f}(\boldsymbol{v})\| \;\leqslant\; (1+\epsilon)\,\|\boldsymbol{u}-\boldsymbol{v}\|,$$

for all $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{S}$.

$\Rightarrow$ isometry between $(\mathcal{S}, \ell_2)$ and $(\boldsymbol{f}(\mathcal{S}), \ell_2)$

Possible mapping: $\boldsymbol{f}(\boldsymbol{u}) = \boldsymbol{\Phi}\boldsymbol{u}$, with, *e.g.*, $\Phi_{ij} \sim_{\rm iid} \mathcal{N}(0,1)$ (or subgaussian)

2. Beyond 1-bit ... Quantizing JL lemma?

First quantization attempt:

Given a JL mapping $\boldsymbol{f} : \mathbb{R}^N \to \mathbb{R}^M$

Form $\boldsymbol{\psi} := \mathcal{Q} \circ \boldsymbol{f}$ with $\mathcal{Q}(t) = \delta(\lfloor t/\delta \rfloor + 1/2)$ (Quantization bin $\delta > 0$)

Then, with $M \geqslant M_0 = O(\epsilon^{-2} \log |\mathcal{S}|)$,

$$\boxed{(1-\epsilon)\,\|\boldsymbol{u}-\boldsymbol{v}\| - \delta \;\leqslant\; \tfrac{1}{\sqrt{M}}\,\|\boldsymbol{\psi}(\boldsymbol{u}) - \boldsymbol{\psi}(\boldsymbol{v})\| \;\leqslant\; (1+\epsilon)\,\|\boldsymbol{u}-\boldsymbol{v}\| + \delta,}$$

2. Beyond 1-bit ... Quantizing JL lemma?

First quantization attempt:

Given a JL mapping $\boldsymbol{f} : \mathbb{R}^N \to \mathbb{R}^M$

Form $\boldsymbol{\psi} := \mathcal{Q} \circ \boldsymbol{f}$ with $\mathcal{Q}(t) = \delta(\lfloor t/\delta \rfloor + 1/2)$ (Quantization bin $\delta > 0$)

Then, with $M \geqslant M_0 = O(\epsilon^{-2} \log |\mathcal{S}|)$,

$$(1-\epsilon)\,\|\boldsymbol{u}-\boldsymbol{v}\| - \delta \;\leqslant\; \tfrac{1}{\sqrt{M}}\,\|\boldsymbol{\psi}(\boldsymbol{u}) - \boldsymbol{\psi}(\boldsymbol{v})\| \;\leqslant\; (1+\epsilon)\,\|\boldsymbol{u}-\boldsymbol{v}\| + \delta,$$

Proof: easy, just observe that:

$$|a-b| - \delta \leq |\mathcal{Q}(a) - \mathcal{Q}(b)| \leq |a-b| + \delta, \quad \forall\, a, b \in \mathbb{R}$$

2. Beyond 1-bit ... Quantizing JL lemma?

First quantization attempt:

Given a JL mapping $\boldsymbol{f}: \mathbb{R}^N \to \mathbb{R}^M$

Form $\boldsymbol{\psi} := \mathcal{Q} \circ \boldsymbol{f}$ with $\mathcal{Q}(t) = \delta(\lfloor t/\delta \rfloor + 1/2)$ (Quantization bin $\delta > 0$)

Then, with $M \geqslant M_0 = O(\epsilon^{-2} \log |\mathcal{S}|)$,

$$\boxed{(1-\epsilon)\,\|\boldsymbol{u}-\boldsymbol{v}\|}\ \boxed{-\,\delta} \leqslant \tfrac{1}{\sqrt{M}}\,\|\boldsymbol{\psi}(\boldsymbol{u})-\boldsymbol{\psi}(\boldsymbol{v})\| \leqslant \boxed{(1+\epsilon)\,\|\boldsymbol{u}-\boldsymbol{v}\|}\ \boxed{+\,\delta},$$

(decaying, good!) (constant, weird!?)

multiplicative error **additive error**

Problem: $\epsilon = O\big(\sqrt{\log |\mathcal{S}|/M_0}\,\big)$ but δ is constant!

Can we hope better?

2. Beyond 1-bit … Quantizing JL lemma?

Let $\psi(\boldsymbol{x}) = \mathcal{Q}(\boldsymbol{\varphi} \cdot \boldsymbol{x} + u)$, for $\boldsymbol{x} \in \mathbb{R}^N$

random Gaussian projection

$\boldsymbol{\varphi} \sim \mathcal{N}^{N\times 1}(0,1)$

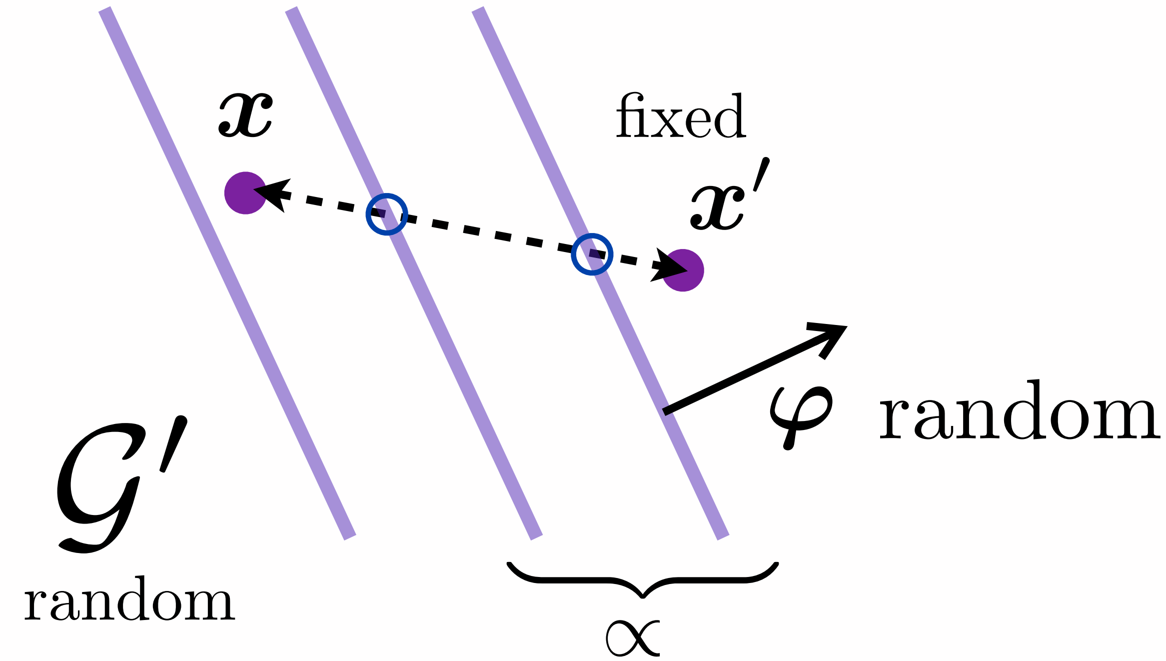
random "dithering"

$u \sim \mathcal{U}([0,\delta])$

Scalar Quantization resolution $\delta > 0$

2. Beyond 1-bit ... Quantizing JL lemma?

Let $\psi(\boldsymbol{x}) = \mathcal{Q}(\boldsymbol{\varphi} \cdot \boldsymbol{x} + u)$, for $\boldsymbol{x} \in \mathbb{R}^N$



$$|\mathcal{Q}(\boldsymbol{\varphi} \cdot \boldsymbol{x} + u) - \mathcal{Q}(\boldsymbol{\varphi} \cdot \boldsymbol{x}' + u)|$$

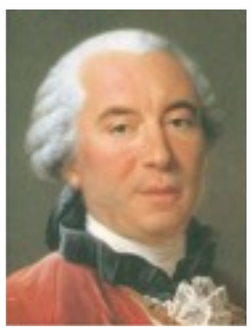
(conditionnally to $\|\boldsymbol{\varphi}\|$)

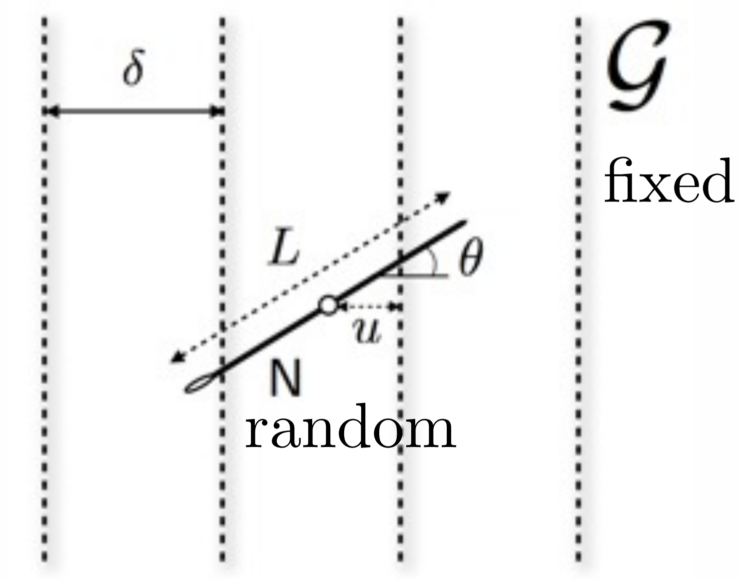
$$=$$

counting planes btw $\boldsymbol{x}$ and $\boldsymbol{x}$' !

2. Beyond 1-bit … Quantizing JL lemma?

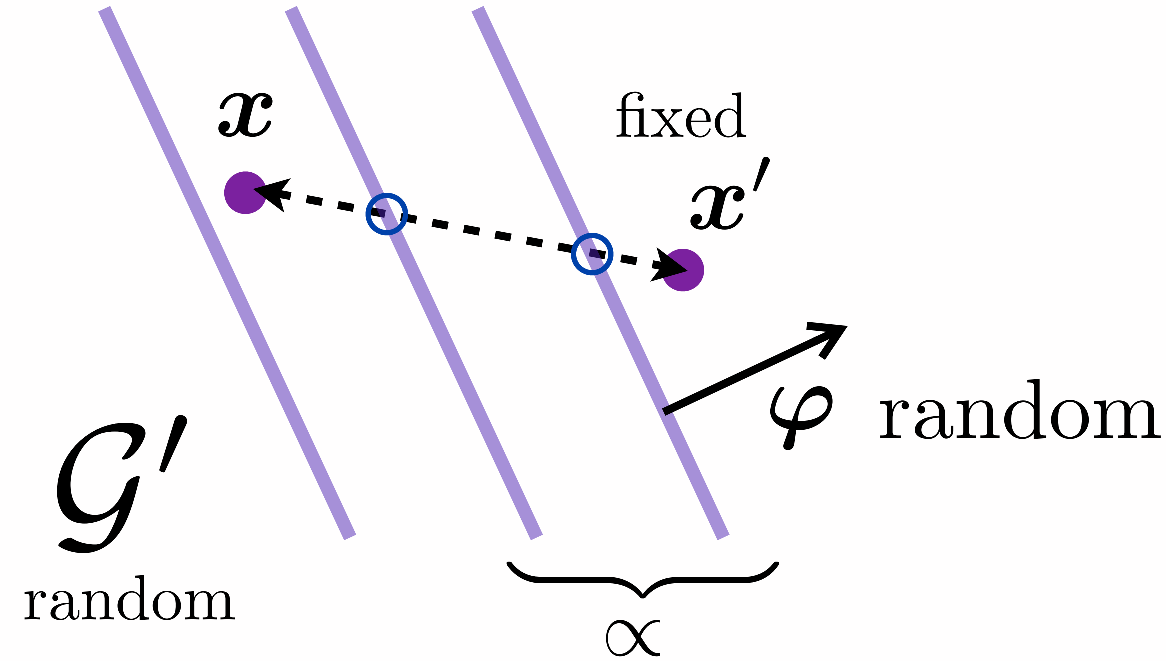
Let $\psi(\boldsymbol{x}) = \mathcal{Q}(\boldsymbol{\varphi} \cdot \boldsymbol{x} + u)$, for $\boldsymbol{x} \in \mathbb{R}^N$





[Buffon's problem 1733, Buffon's solution 1777]

$\equiv$



$|\mathcal{Q}(\boldsymbol{\varphi} \cdot \boldsymbol{x} + u) - \mathcal{Q}(\boldsymbol{\varphi} \cdot \boldsymbol{x}' + u)|$

(conditionnally to $\|\boldsymbol{\varphi}\|$)

Random throw of a "needle" and counts intersections with parallel stripes $\mathcal{G}$.

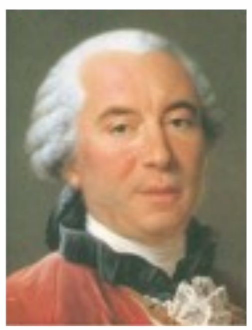
$=$

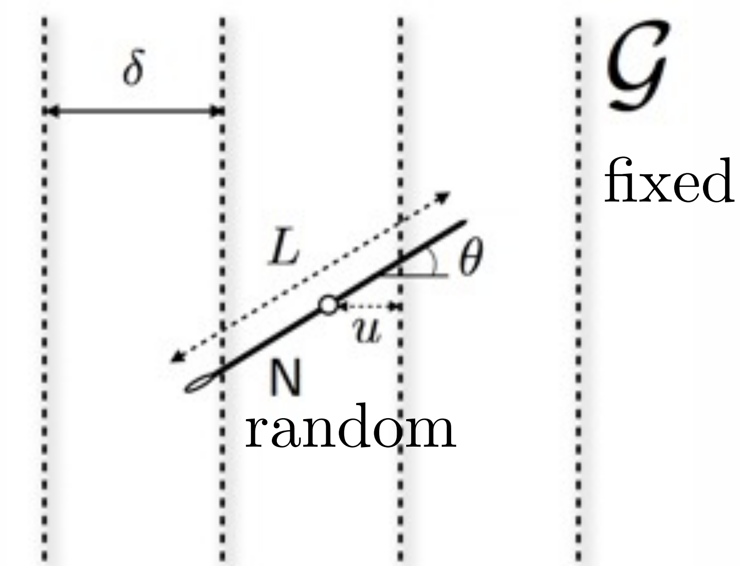
counting planes btw $\boldsymbol{x}$ and $\boldsymbol{x}$' !

$$\mathbb{E}\,|\mathrm{N} \cap \mathcal{G}| = \frac{2}{\pi}\,\frac{L}{\delta}$$

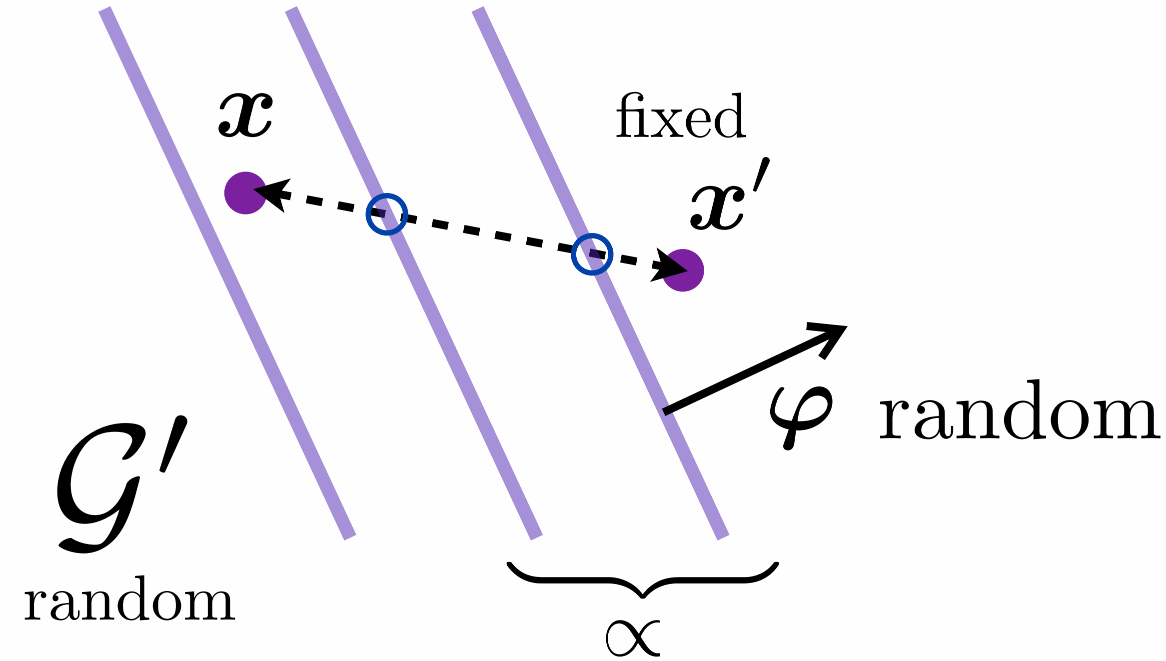
2. Beyond 1-bit … Quantizing JL lemma?

Let $\psi(\boldsymbol{x}) = \mathcal{Q}(\boldsymbol{\varphi} \cdot \boldsymbol{x} + u)$, for $\boldsymbol{x} \in \mathbb{R}^N$





$\equiv$



[Buffon's problem 1733, Buffon's solution 1777]

$$|\mathcal{Q}(\boldsymbol{\varphi} \cdot \boldsymbol{x} + u) - \mathcal{Q}(\boldsymbol{\varphi} \cdot \boldsymbol{x}' + u)|$$

(conditionnally to $\|\boldsymbol{\varphi}\|$)

For M measurements:

$$\boldsymbol{\psi}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{u}) \;\Leftrightarrow\; \psi_j(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\varphi}_j \cdot \boldsymbol{x} + u_j)$$

with: $\boldsymbol{\Phi} \sim \mathcal{N}^{M\times N}(0,1)$
$\boldsymbol{u} \sim \mathcal{U}^M([0,\delta])$

2. Beyond 1-bit … Quantizing JL lemma?

- Let's define the r.v. $X_j = \underline{\frac{1}{\delta}|\psi_j(\boldsymbol{x}) - \psi_j(\boldsymbol{x}')}|$ $\quad (1 \leq j \leq M)$
- Measure concentration for

sub-Gaussian r.v.!

$$\frac{1}{M}\sum_j X_j = \frac{1}{\delta M}\|\boldsymbol{\psi}(\boldsymbol{x}) - \boldsymbol{\psi}(\boldsymbol{x}')\|_1$$

2. Beyond 1-bit ... Quantizing JL lemma?

- Let's define the r.v. $X_j = \frac{1}{\delta}|\psi_j(\boldsymbol{x}) - \psi_j(\boldsymbol{x}')|$ $\quad (1 \leq j \leq M)$
- Measure concentration for

$$\frac{1}{M}\sum_j X_j = \frac{1}{\delta M}\|\boldsymbol{\psi}(\boldsymbol{x}) - \boldsymbol{\psi}(\boldsymbol{x}')\|_1$$

Quasi-isometry! [LJ, 2013]

Lemma 1 *Given an error* $0 < \epsilon < 1$, *and a point set* $\mathcal{S} \subset \mathbb{R}^N$. *If* M *is such that*

$$M \geqslant M_0 = O(\epsilon^{-2}\log|\mathcal{S}|),$$

$$\boxed{\epsilon = O\big(\sqrt{\log|\mathcal{S}|/M_0}\,\big)}$$

then, for $c > 0$ *and with high probability, we have*

$$\boxed{(1-\epsilon)\,\|\boldsymbol{x}-\boldsymbol{x}'\|} - \boxed{c\delta\epsilon} \leqslant \frac{\sqrt{\pi}}{M\sqrt{2}}\,\|\boldsymbol{\psi}(\boldsymbol{x})-\boldsymbol{\psi}(\boldsymbol{x}')\|_1 \leqslant \boxed{(1+\epsilon)\,\|\boldsymbol{x}-\boldsymbol{x}'\|} + \boxed{c\delta\epsilon},$$

for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{S}$.

multiplicative and additive errors decay as $1/\sqrt{M_0}$!

3. Beyond QJL … Quantizing the RIP?

Quantization?

RIP:

$$\boxed{(1-\delta)\|\boldsymbol{u}\|^2 \leq \tfrac{1}{M}\|\boldsymbol{\Phi u}\|^2 \leq (1+\delta)\|\boldsymbol{u}\|^2}$$

for all $\boldsymbol{u} \in \Sigma_{2K} := \{\boldsymbol{u} : \|\boldsymbol{u}\|_0 := |\mathrm{supp}\,\boldsymbol{u}| \leq 2K\}$

- Why quantizing the RIP?
  - since we can ;-)
  - for future algorithm guarantees
  - for nearest neighbors applications

3. Beyond QJL ... Quantizing the RIP?

- Let's retake: for $\mathcal{Q}(\cdot) = \delta\lfloor \cdot/\delta \rfloor \in \delta\mathbb{Z}$

$$\boldsymbol{\psi}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{u}) \;\Leftrightarrow\; \psi_j(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\varphi}_j \cdot \boldsymbol{x} + u_j)$$

with: $\boldsymbol{\Phi} \sim \mathcal{N}^{M\times N}(0,1)$
$\boldsymbol{u} \sim \mathcal{U}^M([0,\delta])$

3. Beyond QJL ... Quantizing the RIP?

- Let's retake: for $\mathcal{Q}(\cdot) = \delta\lfloor \cdot/\delta \rfloor \in \delta\mathbb{Z}$

$$\boldsymbol{\psi}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{u}) \;\Leftrightarrow\; \psi_j(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\varphi}_j \cdot \boldsymbol{x} + u_j)$$

with: $\boldsymbol{\Phi} \sim \mathcal{N}^{M\times N}(0,1)$
$\boldsymbol{u} \sim \mathcal{U}^M([0,\delta])$

YES!

<u>Quantized Gaussian Quasi-Isometric Embedding</u> [LJ, 2015]

Given an error $0 < \epsilon < 1$, *and* $\mathcal{K} \subset \mathbb{R}^N$.
If M *is such that*

$$M \gtrsim \epsilon^{-5} w(\mathcal{K})^2,$$

For $\mathcal{K} = \boldsymbol{A}\Sigma_K \cap \mathbb{B}^N$
and $\boldsymbol{A}$ ONB
$M \gtrsim \epsilon^{-2} K \log \frac{N}{K\delta\epsilon^{3/2}}$

then, for some $c > 0$ *and for all* $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{K}$, *and w.h.p., we have*

$$\Big(\sqrt{\tfrac{2}{\pi}} - \epsilon\Big)\,\|\boldsymbol{x}-\boldsymbol{x}'\| - c\delta\epsilon \;\le\; \tfrac{1}{M}\,\|\boldsymbol{\psi}(\boldsymbol{x}) - \boldsymbol{\psi}(\boldsymbol{x}')\|_1 \;\le\; \Big(\sqrt{\tfrac{2}{\pi}} + \epsilon\Big)\,\|\boldsymbol{x}-\boldsymbol{x}'\| + c\delta\epsilon,$$

3. Beyond QJL ... Quantizing the RIP?

- Let's retake: for $\mathcal{Q}(\cdot) = \delta\lfloor \cdot/\delta \rfloor \in \delta\mathbb{Z}$

$$\boldsymbol{\psi}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{u}) \Leftrightarrow \psi_j(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\varphi}_j \cdot \boldsymbol{x} + u_j)$$

with: $\cancel{\boldsymbol{\Phi} \sim \mathcal{N}^{M \times N}(0,1)}$
$\boldsymbol{u} \sim \mathcal{U}^M([0,\delta])$

OK for sub-Gaussian?
(e.g., Bernoulli)

3. Beyond QJL ... Quantizing the RIP?

- Let's retake: for $\mathcal{Q}(\cdot) = \delta\lfloor \cdot/\delta \rfloor \in \delta\mathbb{Z}$

$$\boldsymbol{\psi}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{u}) \;\Leftrightarrow\; \psi_j(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\varphi}_j \cdot \boldsymbol{x} + u_j)$$

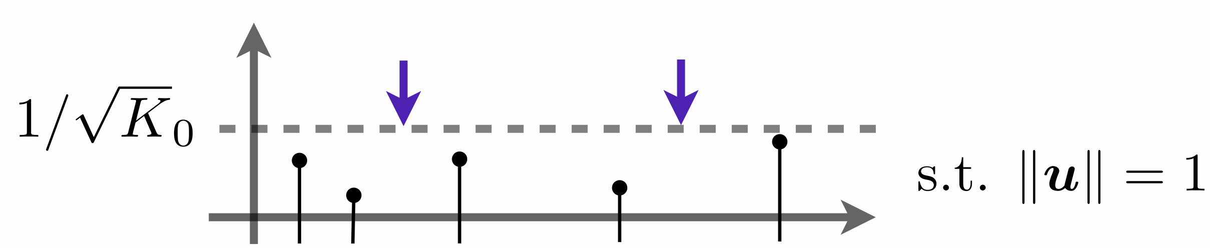
with: $\cancel{\boldsymbol{\Phi} \sim \mathcal{N}^{M\times N}(0,1)}$
$\boldsymbol{u} \sim \mathcal{U}^M([0,\delta])$

OK for sub-Gaussian?
(e.g., Bernoulli)

- **Possible** but:
  - additional distortion! (related to sub-Gaussian dist.)
  - depends on "anti-sparse" nature of $\boldsymbol{x}$ - $\boldsymbol{x}$', *i.e.*,

$$\boldsymbol{x} - \boldsymbol{x}' \in C_{K_0} = \{\boldsymbol{u} \in \mathbb{R}^N : K_0\|\boldsymbol{u}\|_\infty^2 \le \|\boldsymbol{u}\|^2\}$$

for some $K_0 > 0$



3. Beyond QJL ... Quantizing the RIP?

- Let's retake: for $\mathcal{Q}(\cdot) = \delta\lfloor \cdot/\delta \rfloor \in \delta\mathbb{Z}$

$$\boldsymbol{\psi}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi x} + \boldsymbol{u}) \Leftrightarrow \psi_j(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\varphi}_j \cdot \boldsymbol{x} + u_j)$$

with: $\cancel{\boldsymbol{\Phi} \sim \mathcal{N}^{M\times N}(0,1)}$
$\boldsymbol{u} \sim \mathcal{U}^M([0,\delta])$

OK for sub-Gaussian?
(e.g., Bernoulli)

Quantized **sub-**Gaussian Quasi-Isometric Embedding [LJ, 2015]

Given an error $0 < \epsilon < 1$, *and* $\mathcal{K} \subset \mathbb{R}^N$.
If M *is such that*

$$M \gtrsim \epsilon^{-5} w(\mathcal{K})^2,$$

then, w.h.p, for some $c > 0$ *and for all* $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{K}$ *with* $\boldsymbol{x} - \boldsymbol{x}' \in C_{K_0}$, *we have*

$$\Big(\sqrt{\tfrac{2}{\pi}} - \epsilon - \tfrac{\kappa}{\sqrt{K_0}}\Big)\|\boldsymbol{x}-\boldsymbol{x}'\| - c\delta\epsilon \;\le\; \tfrac{1}{M}\|\boldsymbol{\psi}(\boldsymbol{x}) - \boldsymbol{\psi}(\boldsymbol{x}')\|_1 \;\le\; \Big(\sqrt{\tfrac{2}{\pi}} + \epsilon + \tfrac{\kappa}{\sqrt{K_0}}\Big)\|\boldsymbol{x}-\boldsymbol{x}'\| + c\delta\epsilon.$$

For $\mathcal{K} = \boldsymbol{A}\Sigma_K \cap \mathbb{B}^N$
and $\boldsymbol{A}$ ONB
$M \gtrsim \epsilon^{-2} K \log \frac{N}{K\delta\epsilon^{3/2}}$

high K_0, less sparse but lower distortion!

3. Beyond QJL ... Quantizing the RIP?

- Let's retake: for $\mathcal{Q}(\cdot) = \delta\lfloor \cdot/\delta \rfloor \in \delta\mathbb{Z}$

$$\boldsymbol{\psi}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{u}) \Leftrightarrow \psi_j(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\varphi}_j \cdot \boldsymbol{x} + u_j)$$

with: $\cancel{\boldsymbol{\Phi} \sim \mathcal{N}^{M\times N}(0,1)}$
$\boldsymbol{u} \sim \mathcal{U}^M([0,\delta])$

OK for sub-Gaussian?
(e.g., Bernoulli)

<u>Quantized **sub-**Gaussian Quasi-Isometric Embedding</u> [LJ, 2015]

Given an error $0 < \epsilon < 1$, *and* $\mathcal{K} \subset \mathbb{R}^N$.
If M *is such that*

$$M \gtrsim \epsilon^{-5} w(\mathcal{K})^2,$$

then, w.h.p, for some $c > 0$ *and for all* $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{K}$ *with* $\boldsymbol{x} - \boldsymbol{x}' \in C_{K_0}$, *we have*

$$\Big(\sqrt{\tfrac{2}{\pi}} - \epsilon - \tfrac{\kappa}{\sqrt{K_0}}\Big)\,\|\boldsymbol{x}-\boldsymbol{x}'\| - c\delta\epsilon \;\leq\; \tfrac{1}{M}\,\|\boldsymbol{\psi}(\boldsymbol{x}) - \boldsymbol{\psi}(\boldsymbol{x}')\|_1 \;\leq\; \Big(\sqrt{\tfrac{2}{\pi}} + \epsilon + \tfrac{\kappa}{\sqrt{K_0}}\Big)\,\|\boldsymbol{x}-\boldsymbol{x}'\| + c\delta\epsilon.$$

For $\mathcal{K} = \boldsymbol{A}\Sigma_K \cap \mathbb{B}^N$
and $\boldsymbol{A}$ ONB
$M \gtrsim \epsilon^{-2} K \log \frac{N}{K\delta\epsilon^{3/2}}$

But, anti-sparsity "adjustable" with $\boldsymbol{A}$ (e.g., Fourier)!

3. Beyond QJL ... Quantizing the RIP?

- Let's retake: for $\mathcal{Q}(\cdot) = \delta\lfloor \cdot/\delta \rfloor \in \delta\mathbb{Z}$

$$\boldsymbol{\psi}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{u}) \Leftrightarrow \psi_j(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\varphi}_j \cdot \boldsymbol{x} + u_j)$$

with: $\cancel{\boldsymbol{\Phi} \sim \mathcal{N}^{M\times N}(0,1)}$
$\boldsymbol{u} \sim \mathcal{U}^M([0,\delta])$

OK for sub-Gaussian?
(e.g., Bernoulli)

- If you're just asking for consistency:

3. Beyond QJL ... Quantizing the RIP?

- Let's retake: for $\mathcal{Q}(\cdot) = \delta\lfloor \cdot/\delta \rfloor \in \delta\mathbb{Z}$

$$\psi(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{u}) \;\Leftrightarrow\; \psi_j(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\varphi}_j \cdot \boldsymbol{x} + u_j)$$

with: $\cancel{\boldsymbol{\Phi} \sim \mathcal{N}^{M\times N}(0,1)}$
$\boldsymbol{u} \sim \mathcal{U}^M([0,\delta])$

OK for sub-Gaussian?
(e.g., Bernoulli)

- If you're just asking for consistency:

For $\mathcal{K} = \boldsymbol{A}\Sigma_K \cap \mathbb{B}^N$
and $\boldsymbol{A}$ ONB
$M \gtrsim \epsilon^{-1} K \log \frac{N}{K\delta\epsilon^{3/2}}$

Given an error $0 < \epsilon < 1$, *and* $\mathcal{K} \subset \mathbb{R}^N$. *If*

$$M \gtrsim \epsilon^{-4} w(\mathcal{K})^2 \text{ and } \sqrt{K_0} \geq 16\kappa,$$

then, w.h.p., for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{K}$ *with* $\boldsymbol{x} - \boldsymbol{x}' \in C_{K_0}$, *we have*

$$\psi(\boldsymbol{x}) = \psi(\boldsymbol{x}') \quad \Rightarrow \quad \|\boldsymbol{x} - \boldsymbol{x}'\| \leq \epsilon.$$

To conclude ...

Take away messages

- Associating CS and Quantization provides many interesting questions:
  - geometrically (high dim. convex geom.)
  - numerically (not totally covered here)
  - with impacts in CS sensor design

Take away messages

- Associating CS and Quantization provides many interesting questions:
  - geometrically (high dim. convex geom.)
  - numerically (not totally covered here)
  - with impacts in CS sensor design
- Beyond CS, quantifying random projections
  - leads to interesting embedding problems
  - possible impacts in dimensionality reductions

Open questions

- $\ell_2 - \ell_2$ quasi-isometric embedding?
- Embeddings with other quantizations?
- Classification/clustering in the quantized domain?

Thank you for the invitation!

P. T. Boufounos, LJ, F. Krahmer and R. Saab, "**Quantization and Compressive Sensing**", arXiv: 1405.1194, 2014 (to appear in Springer book "Compressed Sensing and Its Applications")

LJ, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "**Robust 1-Bit Compressive Sensing via Binary Stable Embeddings of Sparse Vectors**", IEEE TIT, 59(4), pp. 2082-2102, 2013.

A. Powell, J. Whitehouse, "**Error bounds for consistent reconstruction: random polytopes and coverage processes**", to appear in FoCM, arXiv: 1405.7094

S. Dirksen, G. Lecué, H. Rauhut, "**On the gap between RIP-properties and sparse recovery conditions**" , arXiv: 1504.05073

LJ, "**Error Decay of (almost) Consistent Signal Estimations from Quantized Random Gaussian Projections**", submitted to TIT, arXiv: 1406.0022

LJ, "**A Quantized Johnson Lindenstrauss Lemma: The Finding of Buffon's Needle**", submitted to TIT, arXiv: 1309.1507

LJ, "**Small width, low distortions: quasi-isometric embeddings with quantized sub-Gaussian random projections**", Submitted to TIT, arXiv: 1504.06170

+ references inside the presentation