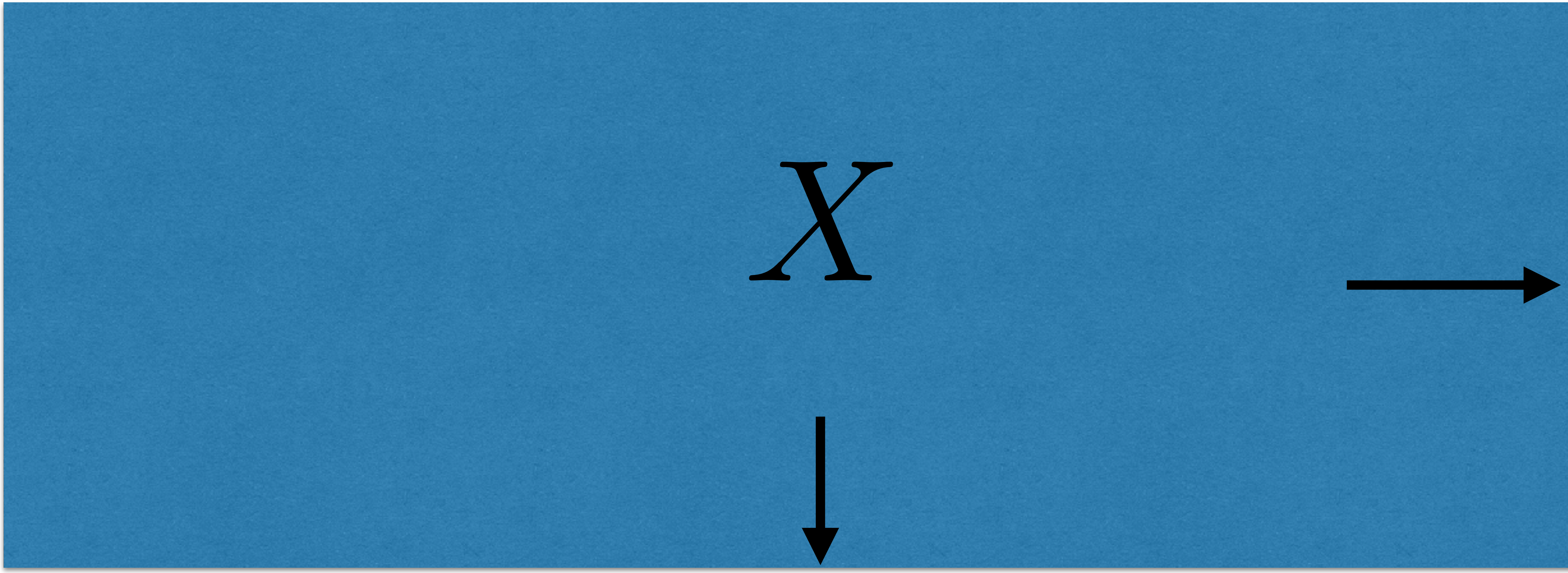
Machine Learning for Data Science (CS4786)
Lecture 6

Compressed Sensing

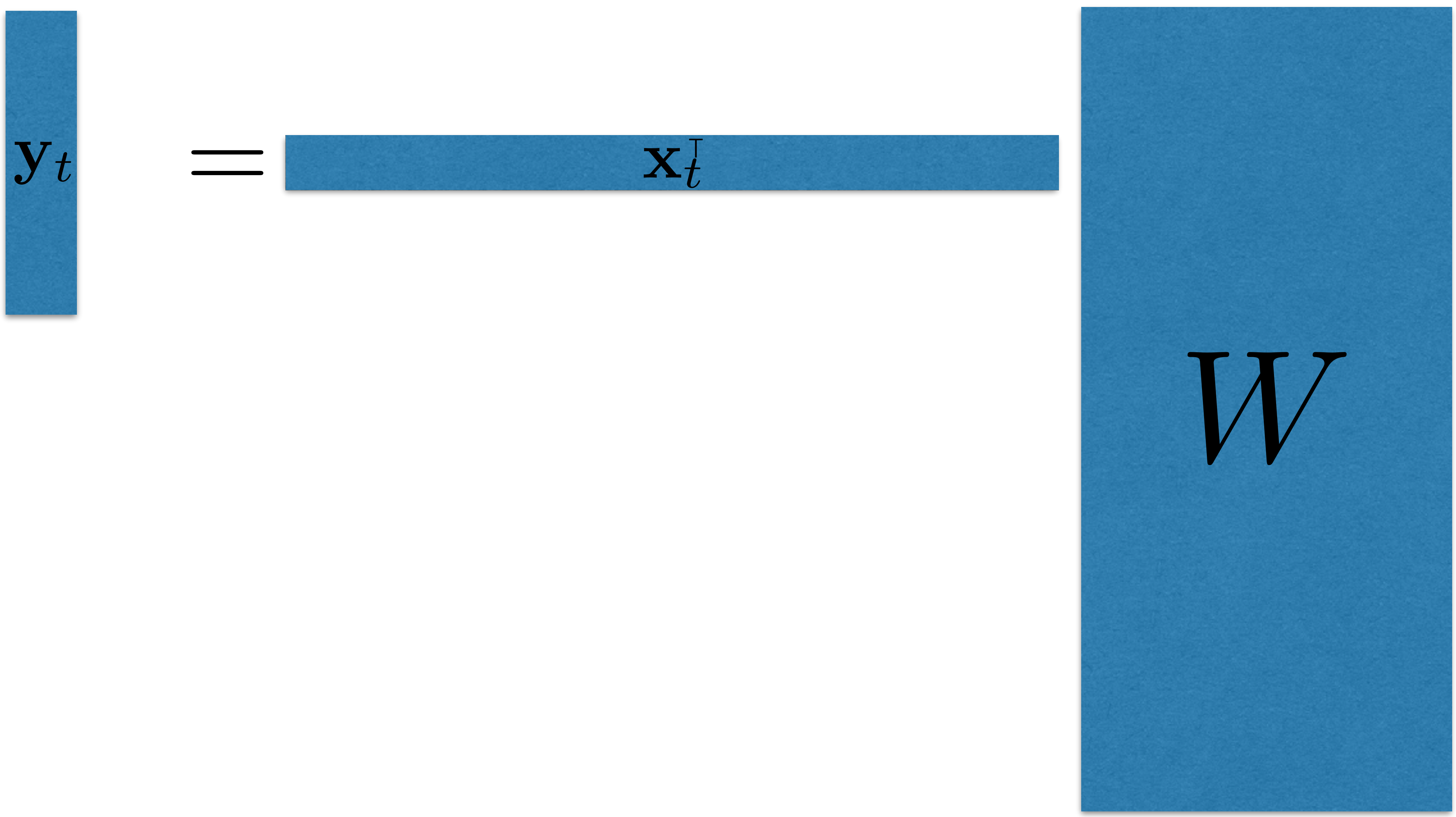
Feb 18, 2016

Course Webpage :
http://www.cs.cornell.edu/Courses/cs4786/2016sp/

THE TALL, THE FAT AND **the Ugly**



- d and n so large we can't even store in memory
- Only have time to be linear in $\text{size}(X) = n \times d$


$\mathbf{y}_t$
$=$
$\mathbf{x}_t^\top$
W

Pick a Random W

$$Y = X \times \left[\begin{array}{ccc} +1 & \ldots & -1 \\ -1 & \ldots & +1 \\ +1 & \ldots & -1 \\ & \vdots & \\ +1 & \ldots & -1 \end{array} \right]_{K}^{d} \Bigg/ \sqrt{K}$$

Random Projections

JL Lemma:
For any $\epsilon > 0$, for K large enough, with high probability over draw of W, for all pairs of data points $i, j \in \{1, \ldots, n\}$,

$$(1 - \epsilon) \left\| \mathbf{y}_i - \mathbf{y}_j \right\|_2 \leq \left\| \mathbf{x}_i - \mathbf{x}_j \right\|_2 \leq (1 + \epsilon) \left\| \mathbf{y}_i - \mathbf{y}_j \right\|_2$$

$$K \approx \frac{\log(n)}{\epsilon^2}$$

Can we always recover $\mathbf{x}_t$'s form $\mathbf{y}_t$'s?

Answer: In general no. When $d > n$ we have an underdetermined system of linear equations.

Can we always recover $\mathbf{x}_t$'s form $\mathbf{y}_t$'s if $\mathbf{x}_t$'s are sparse?

Answer: Yes!

Sparse Data-points

ℓ_0 (norm) of a vector $\mathbf{x} \in \mathbb{R}^d$ measures its "sparsity" and is given by

$$\|\mathbf{x}\|_0 = \#\text{ non-zero entries of } \mathbf{x}$$

Examples:

Recovery for Sparse Data

- When $\mathbf{x}_t$'s are sparse, recovery is possible through random projections.
- Random matrix transformations preserve distances of all sparse vectors!
- This is referred to as restricted isometry property.
- With this property one can successfully perform sparse recovery

Restricted Isometry Property

A projection matrix W of size $K \times d$ possesses (ϵ, s)-RIP, if for all pairs of $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$ with $\|\mathbf{x}\|_0, \|\mathbf{x}'\|_0 \leq s$,

$$(1-\epsilon)\left\|\mathbf{y}-\mathbf{y}'\right\|_2 \leq \left\|\mathbf{x}-\mathbf{x}'\right\|_2 \leq (1+\epsilon)\left\|\mathbf{y}-\mathbf{y}'\right\|_2$$

where $\mathbf{y} = \mathbf{x}^\top W$ and $\mathbf{y}' = \mathbf{x}'^\top W$.

- When $K > \frac{s \log d}{\epsilon^2}$, random matrix W satisfies (ϵ, s)-RIP with high probability.

RIP Implies Sparse Recovery

Algorithm for Recovery:

$$\tilde{\mathbf{x}}_t = \operatorname*{argmin}_{\mathbf{x}:\mathbf{y}_t = \mathbf{x}^\top W} \|\mathbf{x}\|_0$$

Recall definition of RIP:

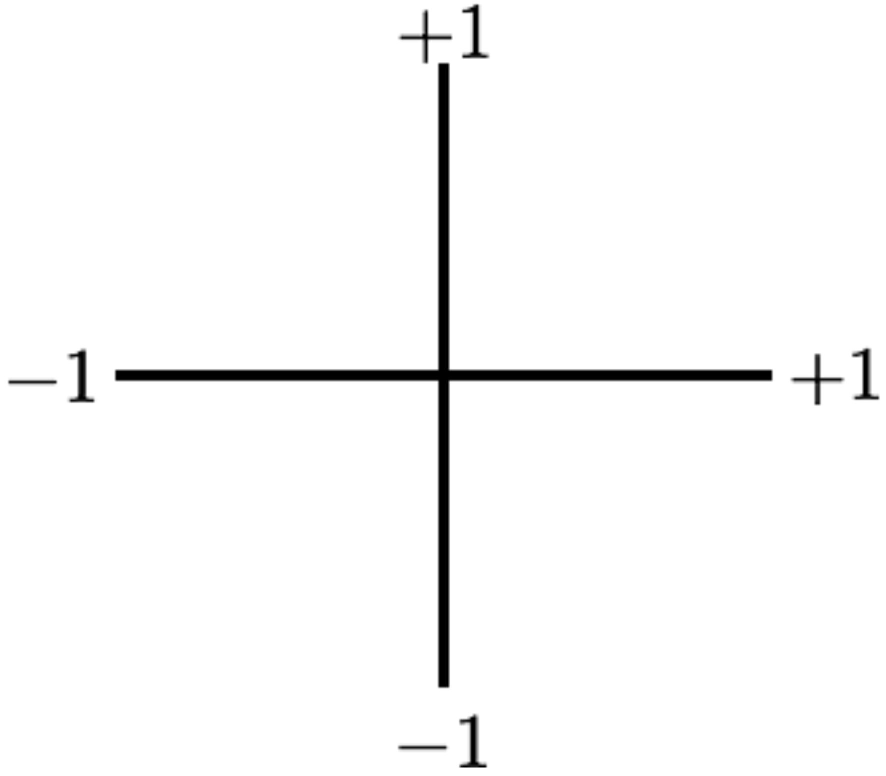
W possesses (ϵ, s)-RIP, if for all pairs of $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$ with $\|\mathbf{x}\|_0, \|\mathbf{x}'\|_0 \leq s$,

$$(1-\epsilon)\|\mathbf{y}-\mathbf{y}'\|_2 \leq \|\mathbf{x}-\mathbf{x}'\|_2 \leq (1+\epsilon)\|\mathbf{y}-\mathbf{y}'\|_2$$

This algorithm is computationally expensive!

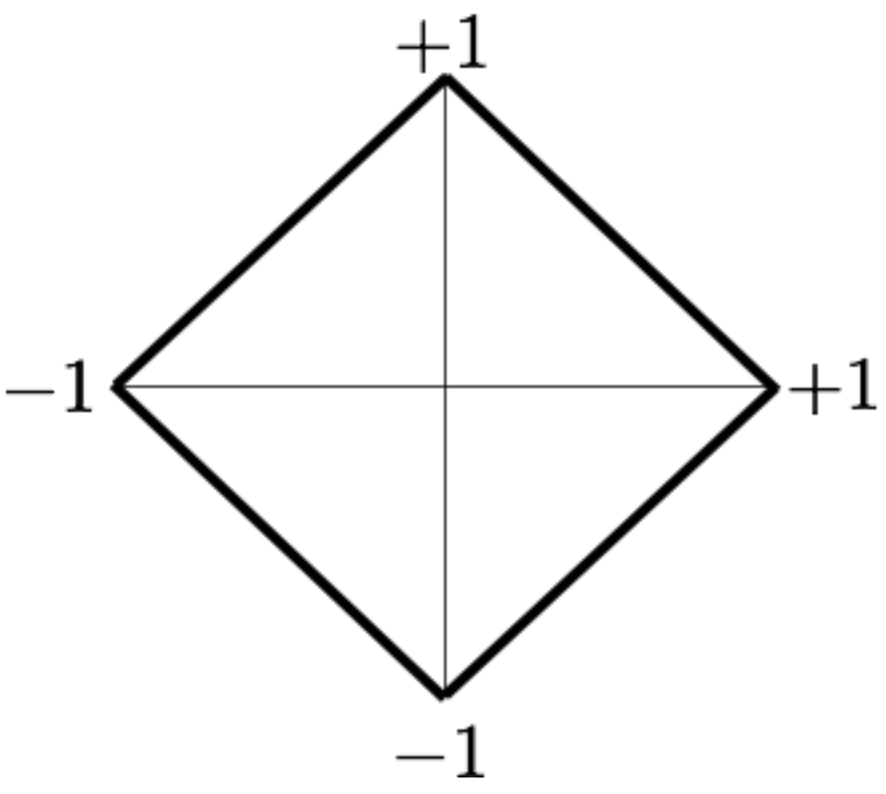
Shape of Sparsity

$$B_0(1) = \left\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_0 \leq 1, \forall i \leq d, |\mathbf{x}[i]| \leq 1\right\}$$



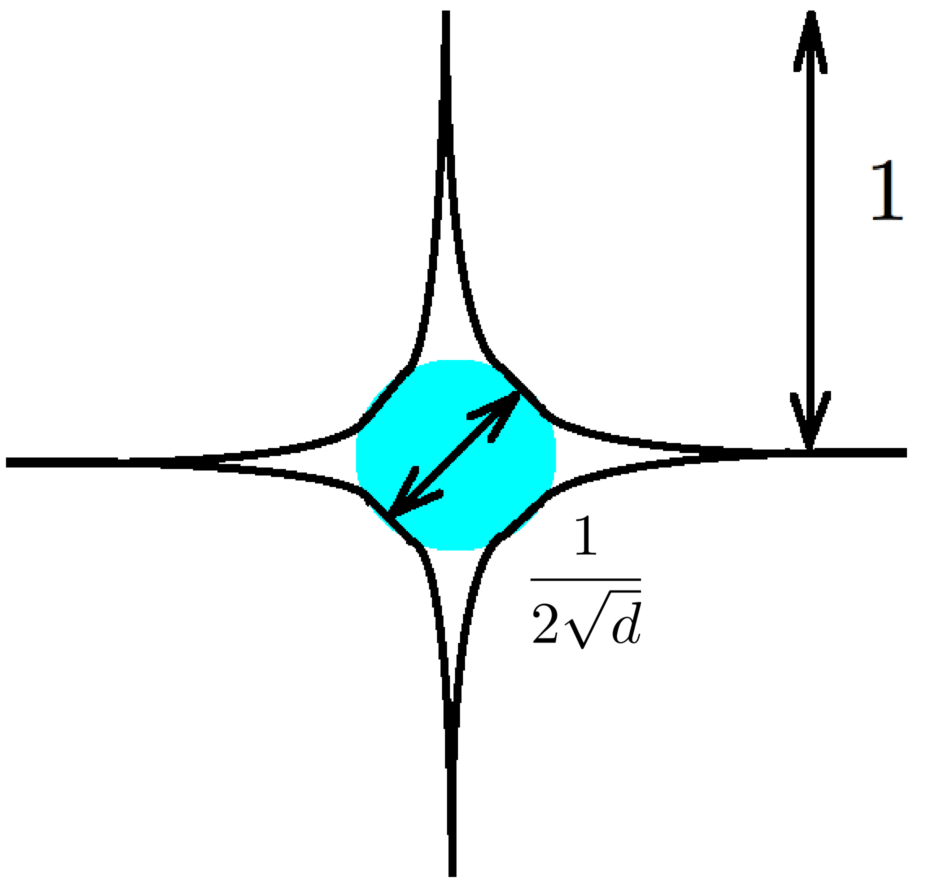
ℓ_1 Ball

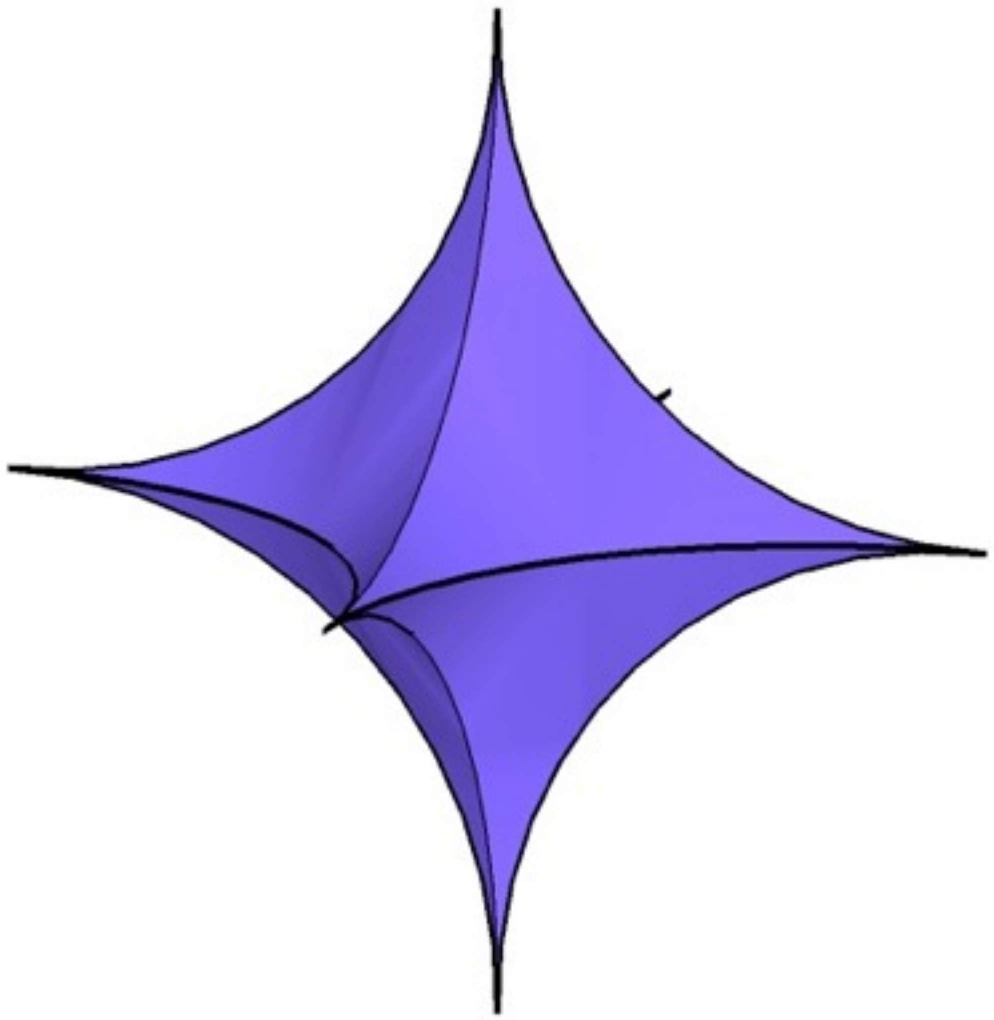
$$B_1(1) = \left\{ \mathbf{x} \in \mathbb{R}^d : \sum_{i=1}^{d} |\mathbf{x}[i]| \leq 1 \right\}$$



ℓ_1 Ball in High Dimensions

Most volume in the center with protruding tentacles reaching out.





$$\|\mathbf{x}\|_1 = \sum_{i=1}^{d} |\mathbf{x}[i]|$$

Replace ℓ_0 by ℓ_1.

Compressed Sensing

1. Perform random projections with large enough K
2. For recovery compute the following:

$$\tilde{\mathbf{x}}_t = \underset{\mathbf{x}:\mathbf{y}_t=\mathbf{x}^\top W}{\operatorname{argmin}} \|\mathbf{x}\|_1$$

   This can be computed efficiently: linear programming problem
3. With high probability for all t's, $\tilde{\mathbf{x}}_t = \mathbf{x}_t$

Compressed Sensing

- If W has (ϵ, s)-RIP then matrix

$$\Phi W$$

has (ϵ', s)-RIP for invertible matrices Φ

- So if data is likely to be sparse under transformation Φ, i.e. $\mathbf{z}_t = \mathbf{x}_t^\top \Phi$ and $\mathbf{z}_t$ is the image we see,
  - Compressed sensing part is the same, Simply project using random projection
  - While reconstructing, use ΦW instead
- Eg. JPEG we use Fourier Transformation, JPEG 2000 Discrete wavelet transformation. If golden standard changes, only minor change in reconstruction, sensing is the same.

Compressed Sensing

- Used for image compression, instead of capturing image in large file and then compressing, directly capture low dimensional representation through random transform
- Allows fast sensing of signals without processing delays
- Random projection can be pushed to hardware level
- JPEG, JPEG 2000 techniques can be applied during sparse recovery.