Universality Laws in Geometric Random Matrix Theory

Joel A. Tropp

Computing + Mathematical Sciences
California Institute of Technology

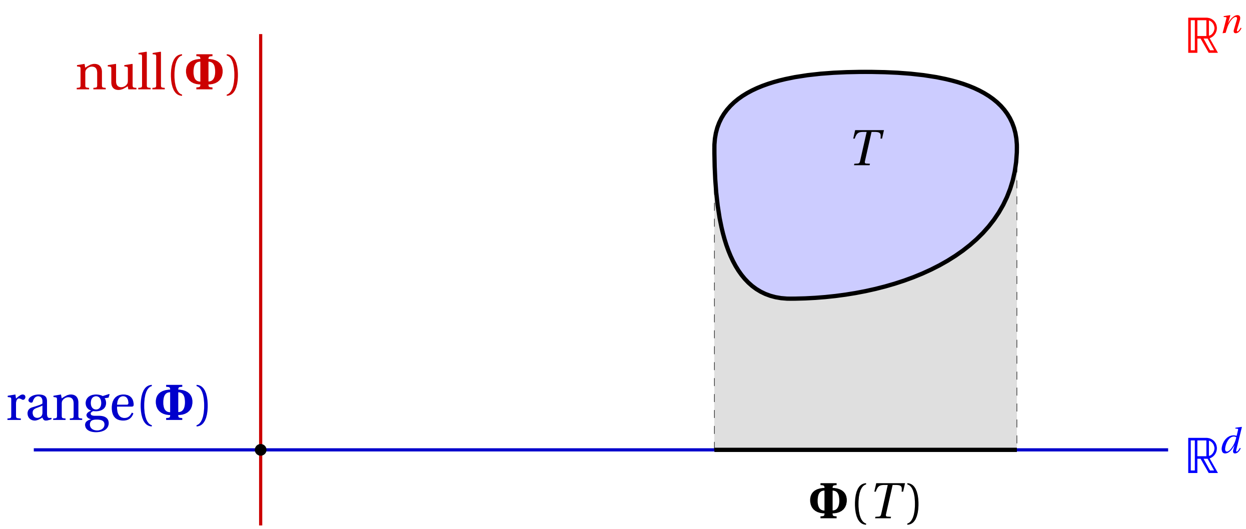
Joint with Samet Oymak (UC-Riverside)

Research supported in part by ONR and the Moore Foundation

Dimension Reduction

Dimension reduction maps a set into lower dimensions,

while preserving features of the geometry



$\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^d$ is a linear dimension reduction map

Dimension Reduction

Dimension reduction maps a set into lower dimensions,

while preserving features of the geometry

Applications:

- **Signal Processing:** Signal acquisition technologies
- **Statistical Estimation:** Experimental designs
- **Coding Theory:** Linear codes are dual to linear dimension reduction
- **Numerical Analysis:** Linear algebra and optimization algorithms

Dimension Reduction: Technical Setup

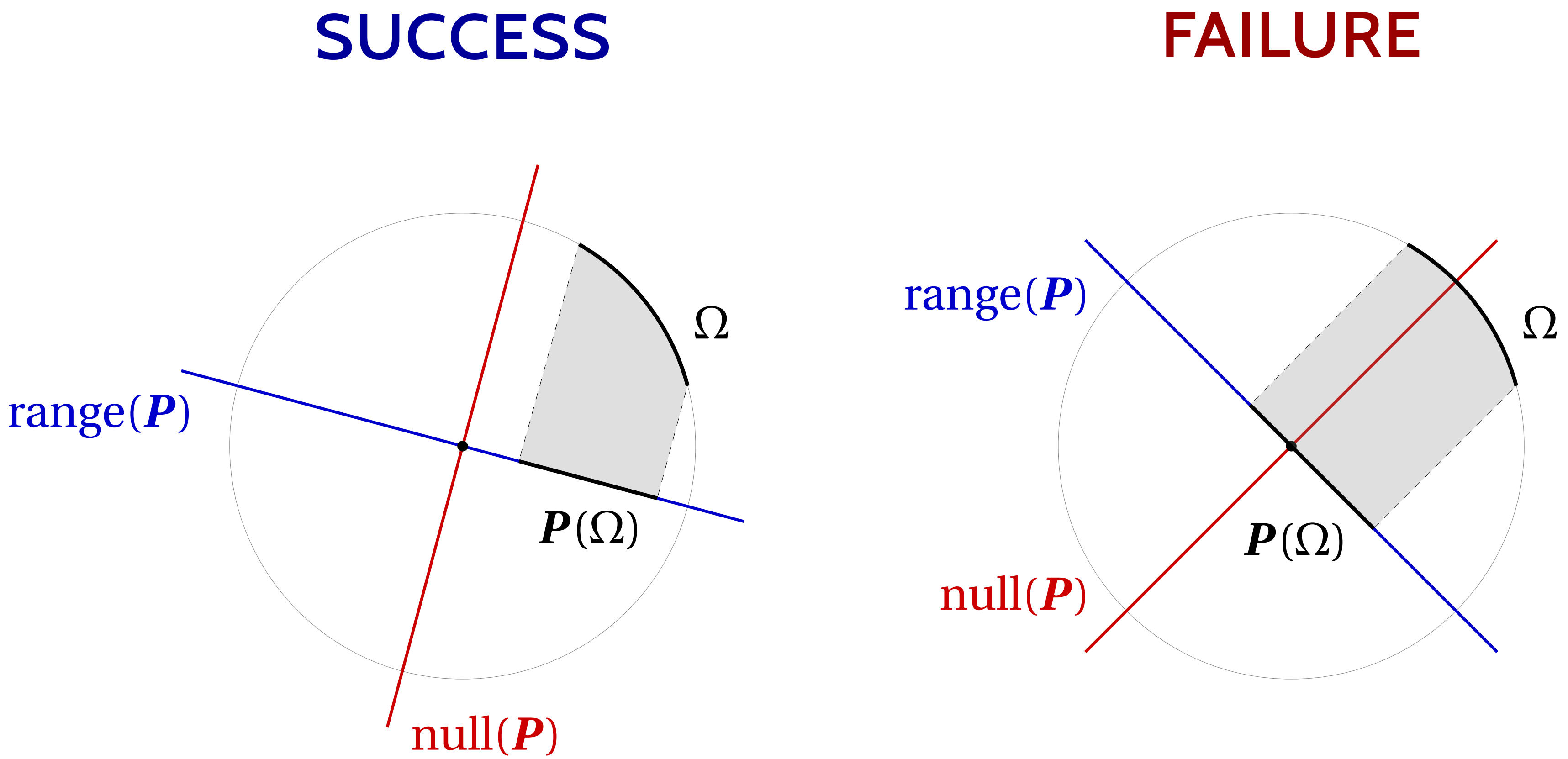
- Let T be a subset of $\mathbb{R}^n$ that does not contain the origin: $\mathbf{0} \notin T$.
- Let $\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^d$ be a linear dimension reduction map ($d \leq n$)
- Want dimension reduction $\mathbf{\Phi}$ to preserve geometric features of T:

SUCCESS: $\mathbf{0} \notin \mathbf{\Phi}(T)$ FAILURE: $\mathbf{0} \in \mathbf{\Phi}(T)$

- Scale-invariant, so pass to the spherical set $\Omega := \{\boldsymbol{t} / \|\boldsymbol{t}\| : \boldsymbol{t} \in T\}$

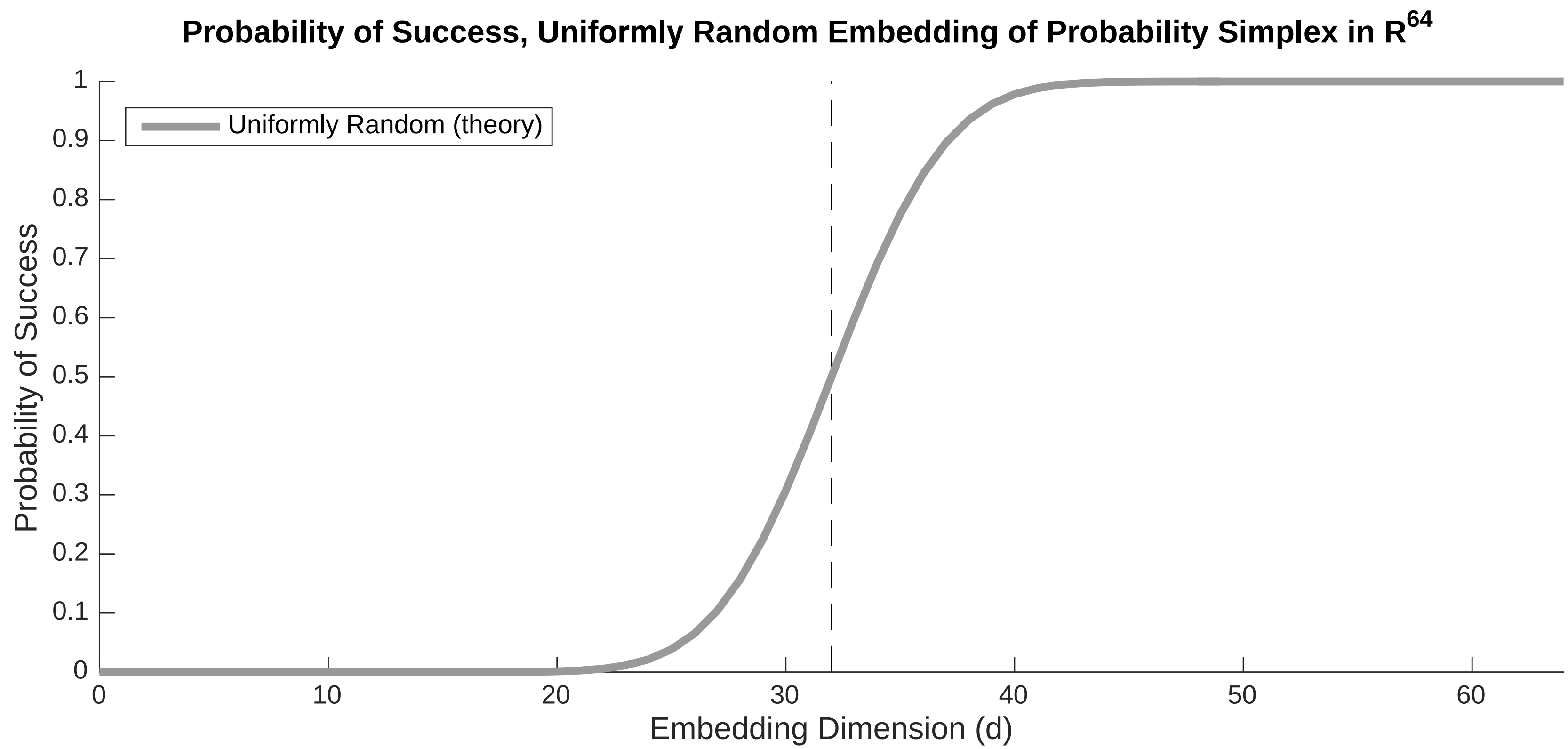
SUCCESS: $\mathbf{0} \notin \mathbf{\Phi}(\Omega)$ FAILURE: $\mathbf{0} \in \mathbf{\Phi}(\Omega)$

Dimension Reduction: Schematic



$\boldsymbol{P} : \mathbb{R}^n \to \mathbb{R}^n$ is the orthogonal projector with $\text{null}(\boldsymbol{P}) = \text{null}(\boldsymbol{\Phi})$

Uniformly Random Embeddings



$$\Delta_{64} = \left\{ \boldsymbol{t} \in \mathbb{R}^{64} : t_i \geq 0, \textstyle\sum_{i=1}^{64} t_i = 1 \right\}$$

$\mathrm{null}(\boldsymbol{\Phi})$ is a uniformly random subspace of $\mathbb{R}^n$ with codimension d

References: Schläfli 1850s; Santaló 1952; Wendell 1962; Cover & Efron 1966; Gordon 1985, 1988; Amelunxen et al. 2014; McCoy & T 2014; Goldstein et al. 2014; Oymak & T 2015; ...

What about Non-Uniform Random Embeddings?

Uniformly random embeddings...

- **(+)** Work very well for applications
- **(+)** Admit precise analysis

- **(–)** May not be implementable
- **(–)** Require expensive construction
- **(–)** May use a lot of storage
- **(–)** Result in expensive arithmetic

What if...

We use discrete, or sparse, or structured random embeddings instead?

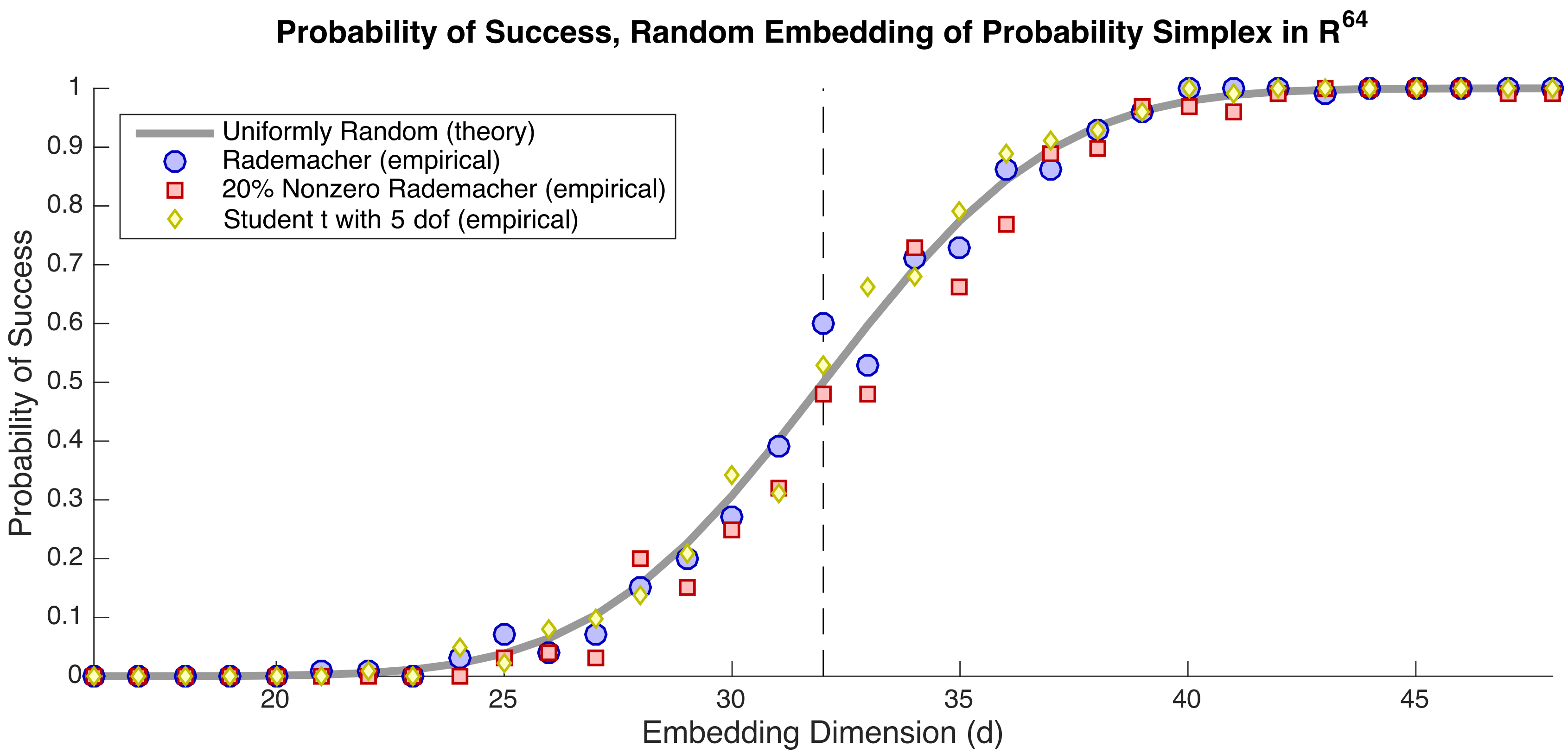
References: Alon et al. 1996; Achlioptas & McSherry 2001, 2007; Guha et al. 2002; Drineas et al. 2004–2006; Martinsson et al. 2006; Sarlós 2006; Ailon & Chazelle 2006, 2009; Candès & Romberg 2007; Woolfe et al. 2007; Liberty 2009; T et al. 2010; Halko et al. 2011; T 2011; Clarkson & Woodruff 2012; Nguyen & Nelson 2012; Boutsidis & Gittens 2012; Gittens & Mahoney 2013; Bourgain et al. 2015;...

Two Random Dimension Reduction Matrices

$$\boldsymbol{\Phi}_{\text{sprad.50}} = \begin{bmatrix} & & +1 & +1 & & & & \\ +1 & -1 & & & & -1 & & +1 \\ -1 & & -1 & -1 & & -1 & & \\ & & +1 & & +1 & & -1 & +1 \end{bmatrix}$$

$$\boldsymbol{\Phi}_{\text{stud5}} = \begin{bmatrix} -1.49 & +1.35 & -0.30 & +0.87 & -0.23 & -0.32 & -0.98 & -0.38 \\ +0.22 & -1.26 & +1.46 & -0.17 & -0.40 & +1.16 & +0.14 & +3.65 \\ -1.05 & -0.12 & -0.31 & -0.81 & -0.43 & -0.41 & +0.79 & +0.70 \\ -1.09 & +0.84 & +1.71 & -1.05 & +0.64 & +1.39 & -0.33 & -0.38 \end{bmatrix}$$

Empirical Behavior of Random Embeddings



$$\Delta_{64} = \left\{ \boldsymbol{t} \in \mathbb{R}^{64} : t_i \geq 0, \sum_{i=1}^{n} t_i = 1 \right\}$$

Success probability seems not to depend on distribution!

Independent Random Embeddings

Model (Independent Random Embedding). Fix a parameter $B \geq 1$.
Let $\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^d$ be a random matrix with the following properties:

- **Independence:** The entries are statistically independent
- **Standardization:** Each entry has mean zero and variance one
- **Symmetry:** Each entry X has the same distribution as its negation $-X$
- **Bounded Moments:** Each entry X satisfies $\mathbb{E}\,|X|^5 \leq B$

Examples...

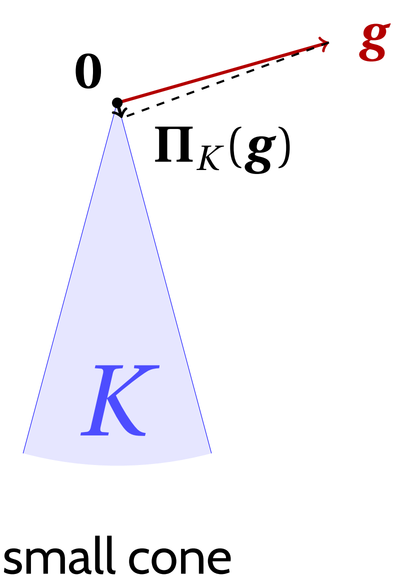
- Gaussian
- Rademacher (± 1)
- Sparse Rademacher
- Student t with 5+ degrees of freedom

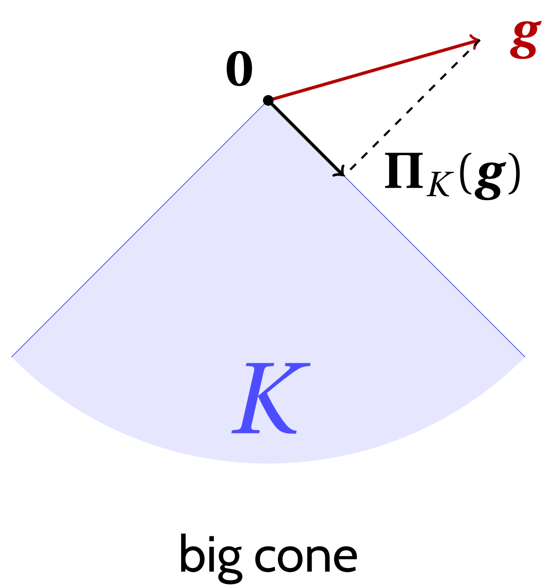
The Statistical Dimension

Definition. Let $T \subset \mathbb{R}^n$. The *statistical dimension* of T is the quantity

$$\delta(T) := \mathbb{E} \left\| \mathbf{\Pi}_{\mathrm{cone}(T)}(\boldsymbol{g}) \right\|_2^2 \quad \text{where} \quad \boldsymbol{g} \sim \textsf{NORMAL}(\mathbf{0}, \mathbf{I}_n)$$

where cone is the conic convex hull and $\mathbf{\Pi}_K$ is the projector onto a convex cone K





The Statistical Dimension

Definition. Let $T \subset \mathbb{R}^n$. The *statistical dimension* of T is the quantity

$$\delta(T) := \mathbb{E}\left\|\mathbf{\Pi}_{\mathrm{cone}(T)}(\boldsymbol{g})\right\|_2^2 \quad \text{where} \quad \boldsymbol{g} \sim \text{NORMAL}(\mathbf{0}, \mathbf{I}_n)$$

where cone is the conic convex hull and $\mathbf{\Pi}_K$ is the projector onto a convex cone K

Basic Properties

- If $E \subset T \subset \mathbb{R}^n$, then $0 \leq \delta(E) \leq \delta(T) \leq n$
- If L is a subspace, then $\delta(L) = \dim(L)$
- If K is a closed convex cone, then $\delta(K^\circ) = n - \delta(K)$
- $\delta(T)$ can be calculated very accurately for many choices of T
- **Example:** $\delta(\Delta_n) = n/2$

References: Gordon 1985, 1988; Rudelson & Vershynin 2006; Stojnic 2009; Oymak & Hassibi 2010; Chandrasekaran et al. 2012; Amelunxen et al. 2014., McCoy & T 2014; Foygel & Mackey 2014; T 2015; Vershynin 2015; ...

A Universality Law for Randomized Embedding

Theorem 1 (Oymak & T 2015). *Suppose that*

- T *is a compact subset of* $\mathbb{R}^n$ *that does not contain the origin*
- $\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^d$ *is an independent random embedding with parameter* B

Then

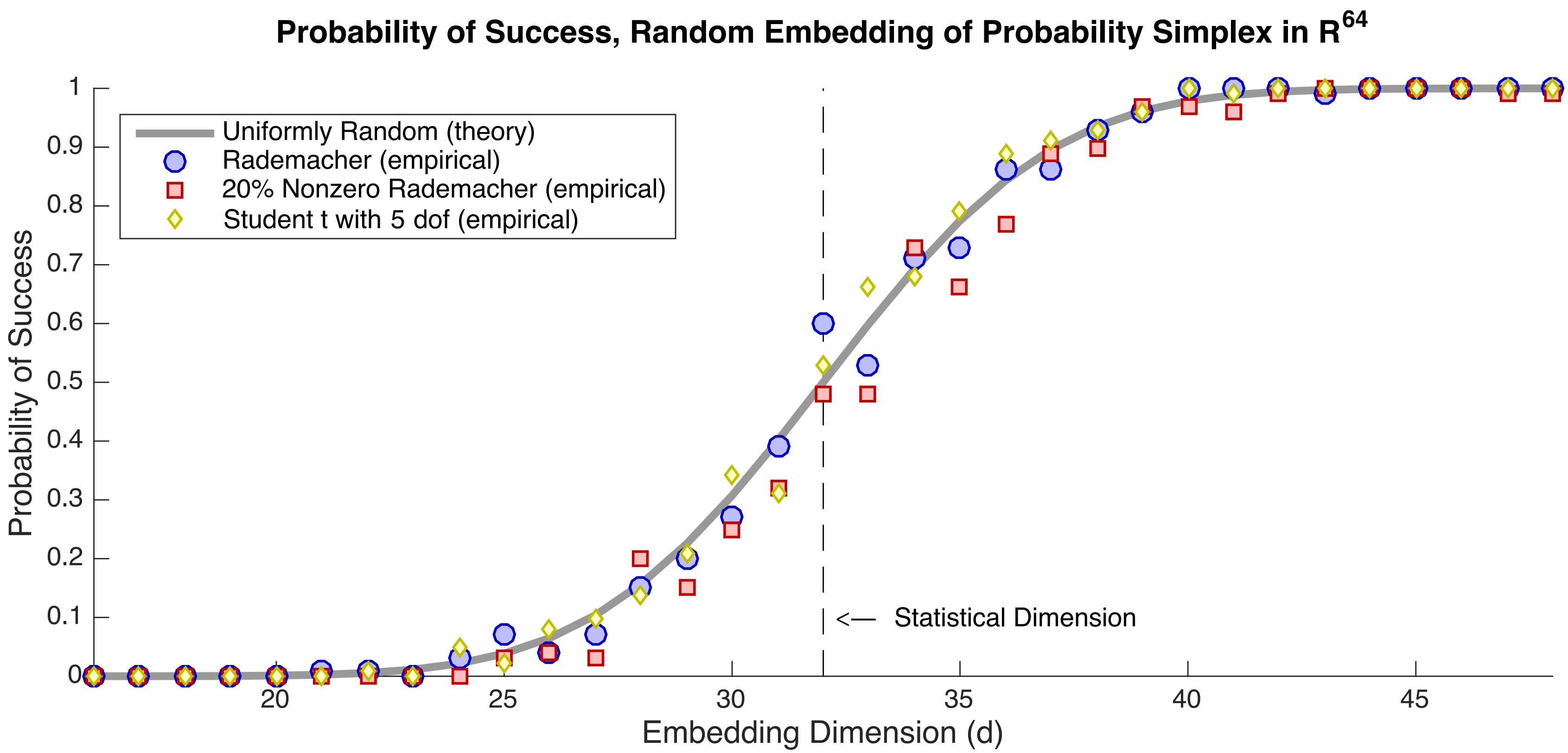
$$d \geq \delta(T) + o(n) \quad \text{implies} \quad \mathbf{0} \notin \mathbf{\Phi}(T) \quad \text{with high prob.} \qquad \text{(SUCCESS)}$$

Furthermore, if the positive hull $\bigcup_{\alpha \geq 0} \alpha T$ *is convex,*

$$d \leq \delta(T) - o(n) \quad \text{implies} \quad \mathbf{0} \in \mathbf{\Phi}(T) \quad \text{with high prob.} \qquad \text{(FAILURE)}$$

The little-o suppresses constants that depend only on the parameter B.

Theoretical Behavior of Random Embeddings



$$\Delta_{64} = \left\{ \boldsymbol{t} \in \mathbb{R}^{64} : t_i \geq 0, \sum_{i=1}^{64} t_i = 1 \right\}$$

Phase transition **does not** depend on distribution!

The Compressed Sensing Problem

Sparse signal acquisition via randomized dimension reduction

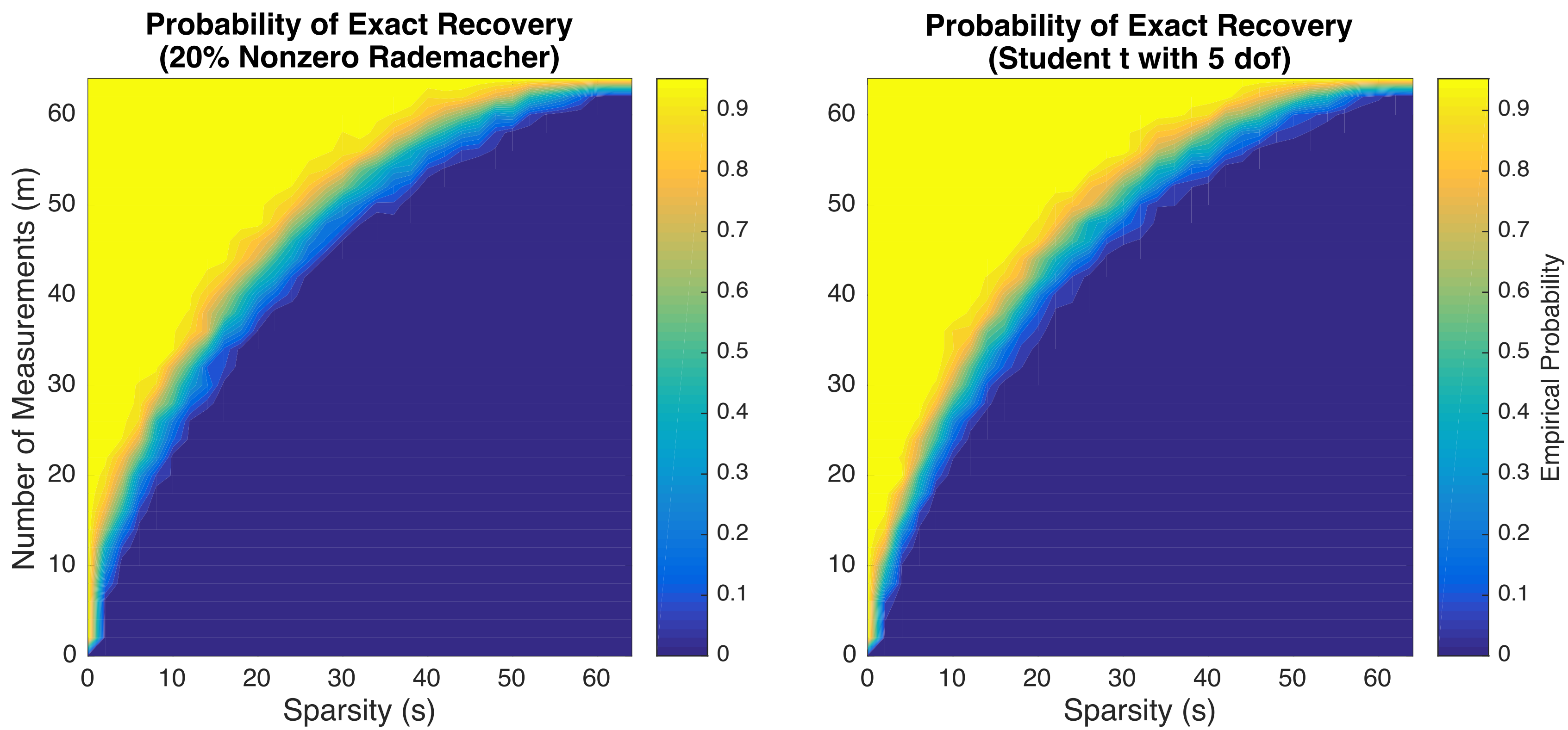
- Let $\boldsymbol{x}^\natural \in \mathbb{R}^n$ be an unknown *sparse* vector with s nonzero entries
- Let $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n}$ be a (random) measurement matrix
- Observe m random measurements: $\boldsymbol{z} = \boldsymbol{\Phi}\boldsymbol{x}^\natural$
- Produce an estimate $\widehat{\boldsymbol{x}}$ by solving the convex program

$$\text{minimize} \quad \|\boldsymbol{x}\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{\Phi}\boldsymbol{x} = \boldsymbol{z}$$

- **Success:** $\widehat{\boldsymbol{x}} = \boldsymbol{x}^\natural$

References: Chen et al. 1996; Donoho & Huo 2001; Fuchs 2006; T 2006; Candès et al. 2006; Donoho 2006; ...

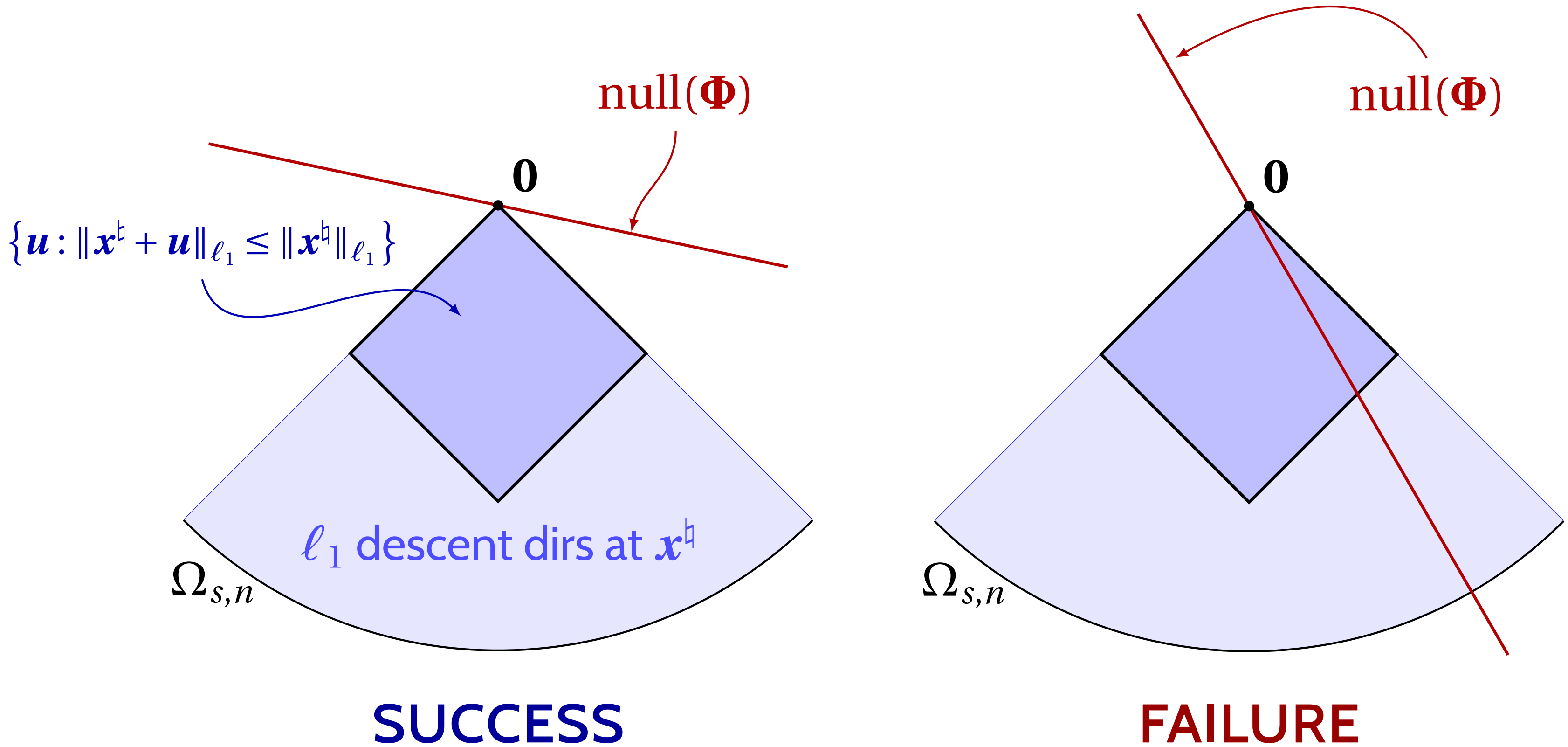
A Computer Experiment and a Mystery



What's going on here?

References: Donoho & Tanner 2009ab; Bayati et al. 2015.

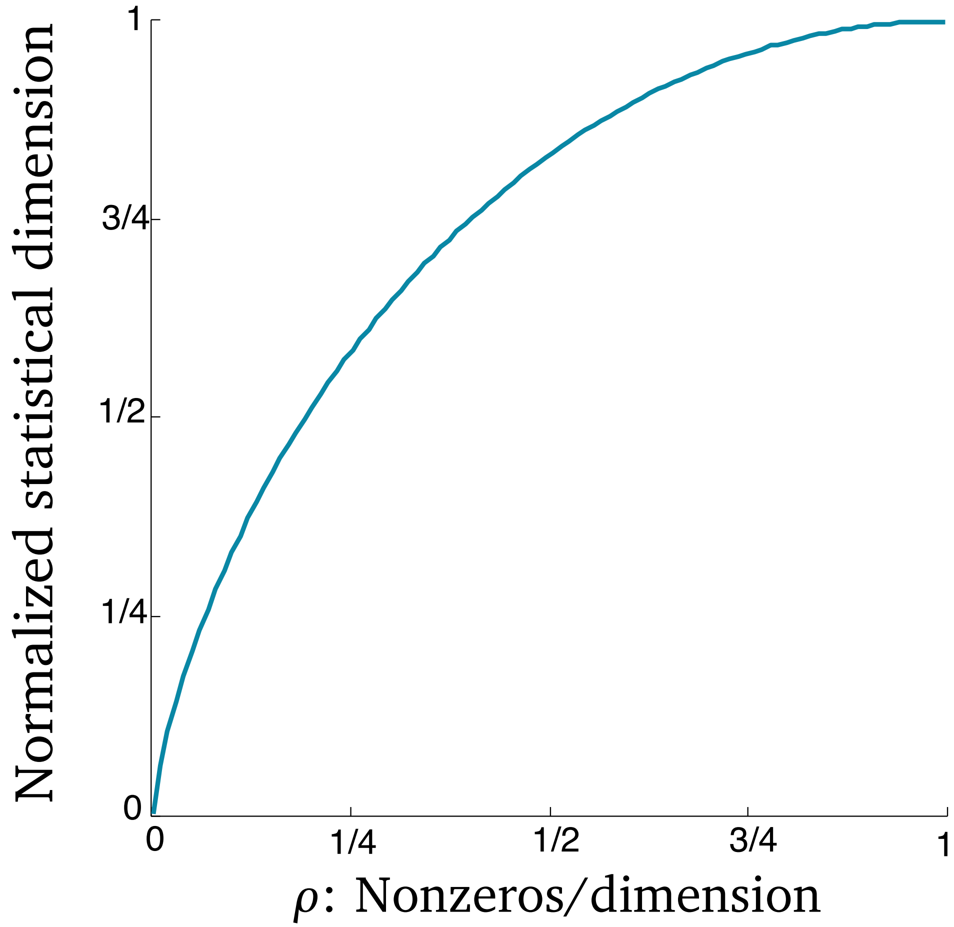
Geometry of Compressed Sensing



$$\text{minimize} \quad \|\boldsymbol{x}^\natural + \boldsymbol{u}\|_{\ell_1} \quad \text{subject to} \quad \mathbf{\Phi}\boldsymbol{u} = \mathbf{0}$$

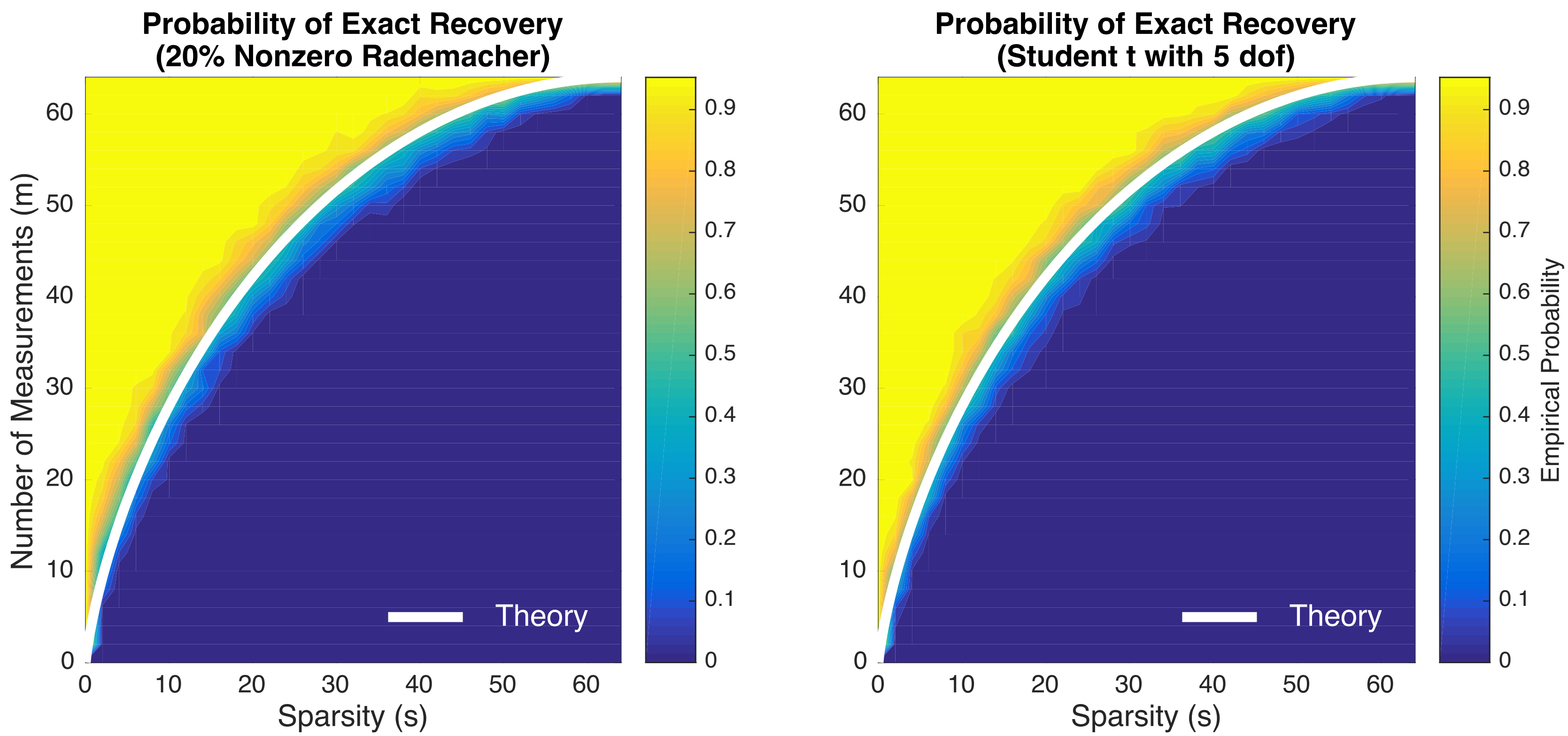
References: Candès et al. 2006; Rudelson & Vershynin 2006, Stojnic 2009; Chandrasekaran et al. 2012; Amelunxen et al. 2014; ...

The Statistical Dimension of $\Omega_{s,n}$



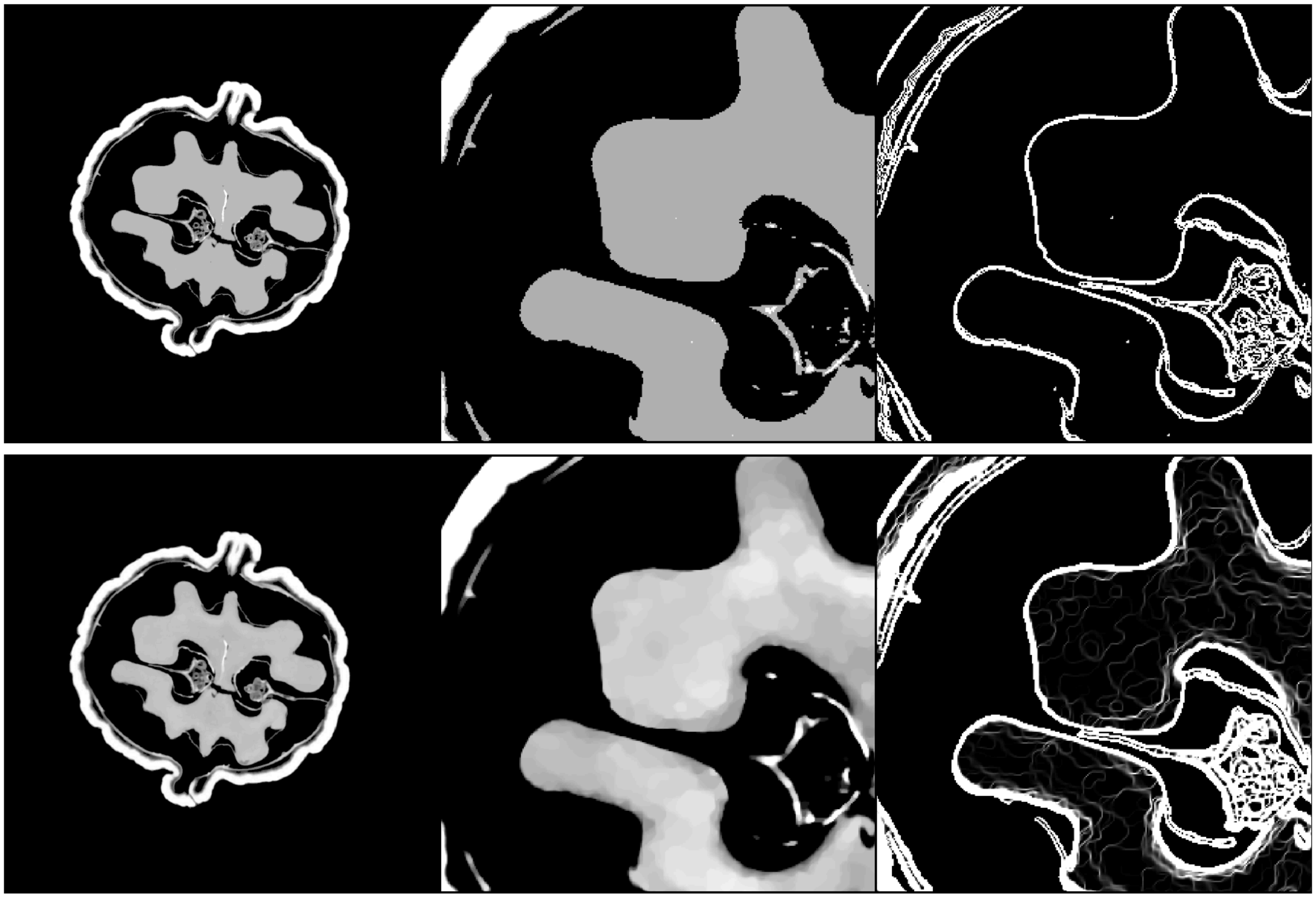
References: Ruben 1960; Vershik & Sporyshev 1986, 1992; Affentranger & Schneider 1992; Betke & Henk 1993; Böröczky & Henk 1999; Donoho & Tanner 2004–2009; Stojnic 2009; Chandrasekaran et al. 2012; Amelunxen et al. 2014; Foygel & Mackey 2014; ...

Universal Behavior in Compressed Sensing



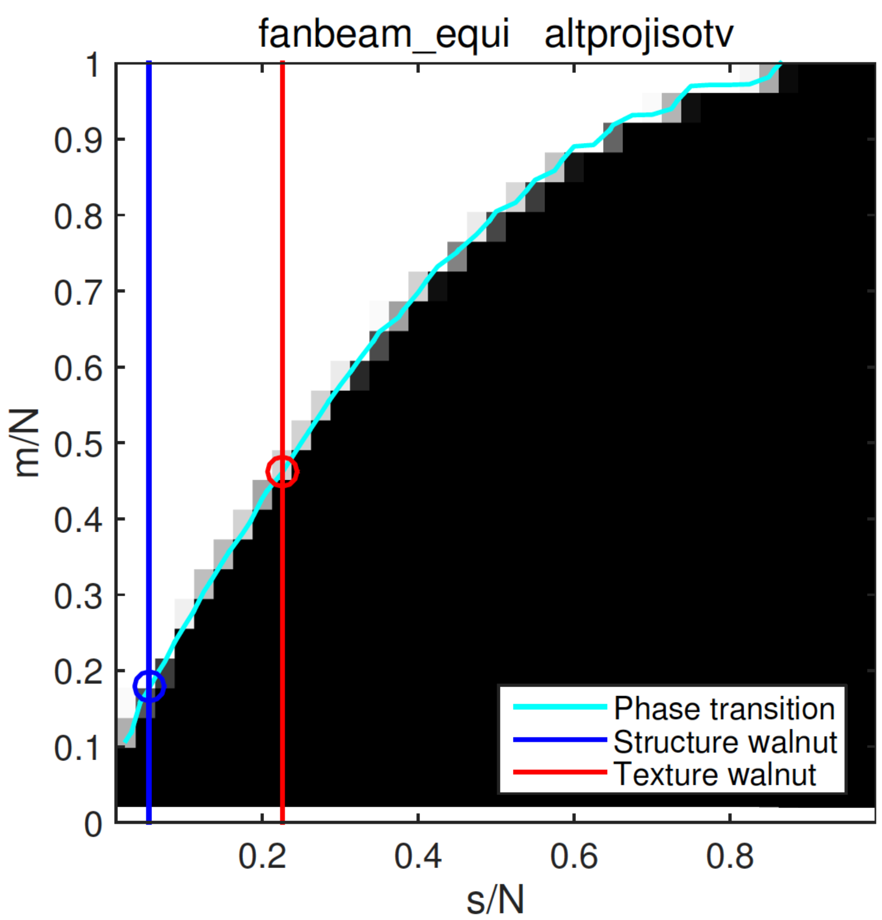
Universality law for random embedding predicts phase transition!

Case Study: Walnut Phantoms



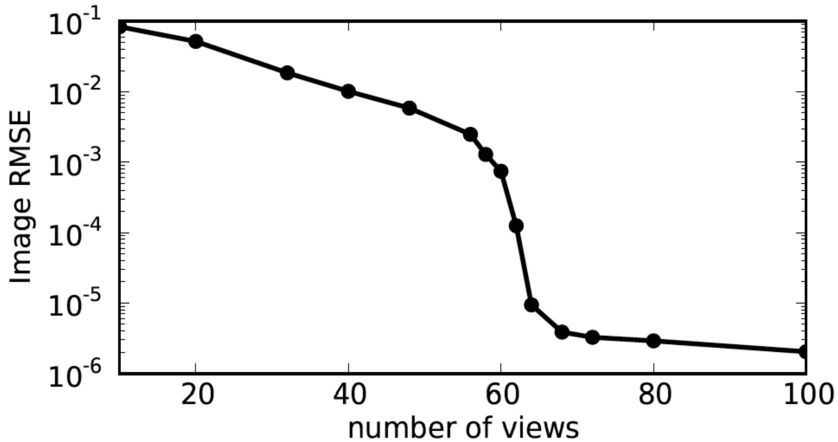
Reference: Jørgensen & Sidky 2014.

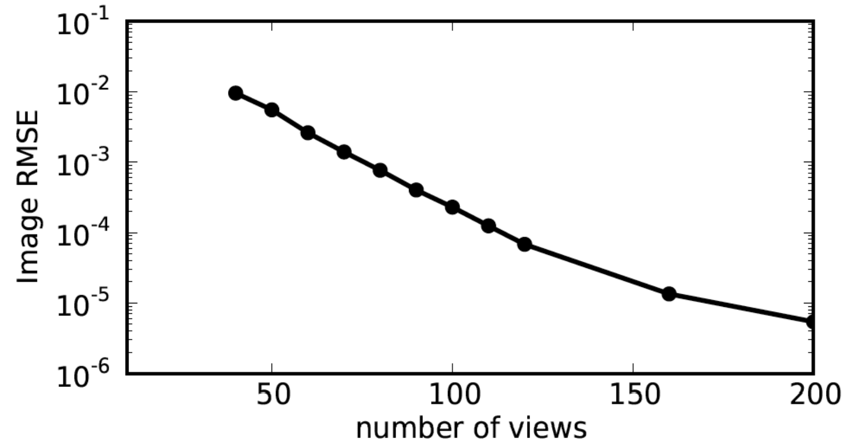
Case Study: Walnut Phantoms



Reference: Jørgensen & Sidky 2014.

Case Study: Walnut Phantoms



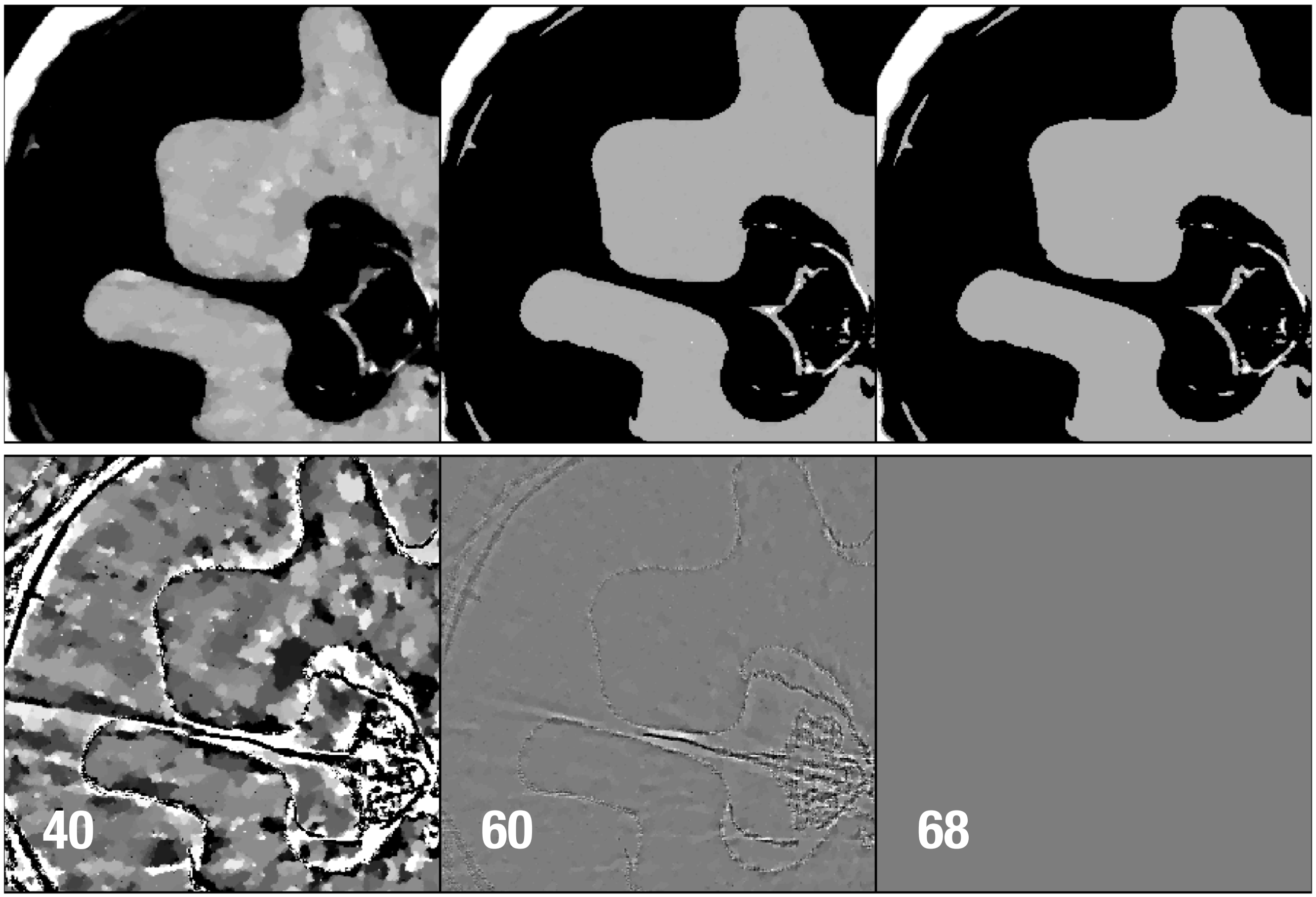


Walnut image	Gradient sparsity	Recovered at	DT prediction	ALMT prediction
Structure	45, 074	68	69.3	71.7
Texture	186, 306	?	188.7	185.8

Table 1: Walnut test images with gradient-domain sparsity levels, number of projections at which recovery is observed, and DT and ALMT phase-diagram predictions of critical sampling levels. A reference point of full sampling is $N_v \geq 403$ projections, where the system matrix has more rows than columns.

Reference: Jørgensen & Sidky 2014.

Case Study: Walnut Phantoms



Reference: Jørgensen & Sidky 2014.

Other Applications...

- **Signal Processing:** Regularized detection, classification, and reconstruction
- **Statistical Estimation:** Regularized least-squares and least absolute deviation
- **Coding Theory:** Error tolerance of random codebooks
- **Numerical Analysis:** Better randomized linear algebra and optimization
- **Stochastic Geometry:** Facial structure of convex hulls of random vectors
- **Random Matrix Theory:** Minimum singular value of a random matrix
- **Neuroscience?!** The brain may perform dimension reduction...

To learn more...

Email: `jtropp@cms.caltech.edu`

Web: `http://users.cms.caltech.edu/~jtropp`

Main Papers Discussed:

- Gordon, "On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$," *GAFA*, 1988
- Rudelson & Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *CPAM*, 2008
- Stojnic, "Various thresholds for ℓ_1 optimization in compressed sensing," arXiv 0907.3666
- Donoho & Tanner, "Observed universality of phase transitions...," *Phil. Trans. Roy. Soc. London*, 2009
- Donoho & Tanner, "Counting faces of randomly projected polytopes...," *JAMS*, 2009
- Chandrasekaran et al., "The convex geometry of linear inverse problems," *FOCM*, 2012
- Amelunxen et al., "Living on the edge: Phase transitions in convex programs with random data," *I&I*, 2014
- Stojnic, "Regularly random duality," arXiv 1303.7295
- Oymak, Thrampoulidis & Hassibi, "The squared error of generalized LASSO: A Precise Analysis," arXiv 1311.0830
- Thrampoulidis, Oymak & Hassibi, "The Gaussian Min–Max Theorem in the Presence of Convexity," arXiv 1408.4837
- Jørgensen & Sidky, "How little data is enough? ...," arXiv 1412.6833

- Oymak & T, "Universality laws for randomized dimension reduction, with applications," *I&I*, 2017