ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Winter Conference in Statistics 2013

Compressed Sensing

LECTURE #11

Compressible priors

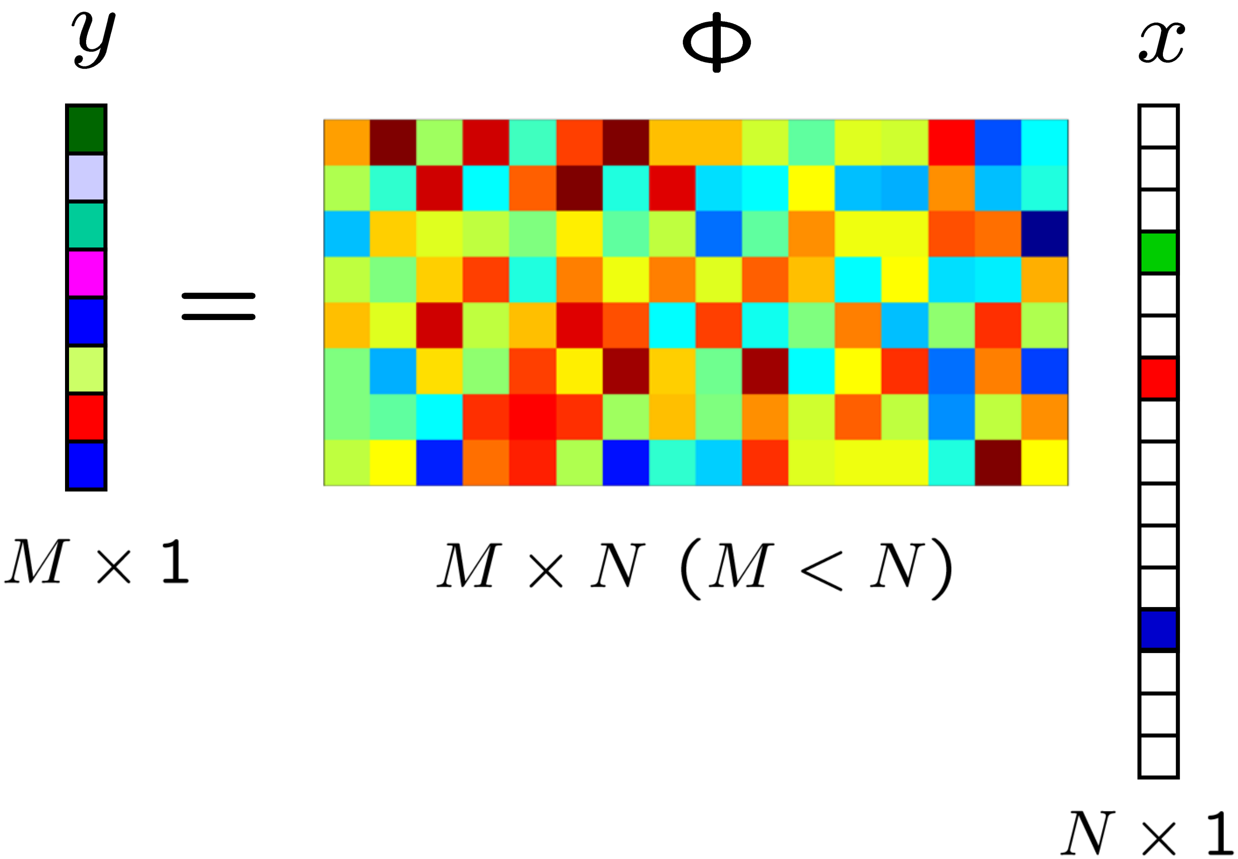
lions@epfl

Prof. Dr. Volkan Cevher

volkan.cevher@epfl.ch

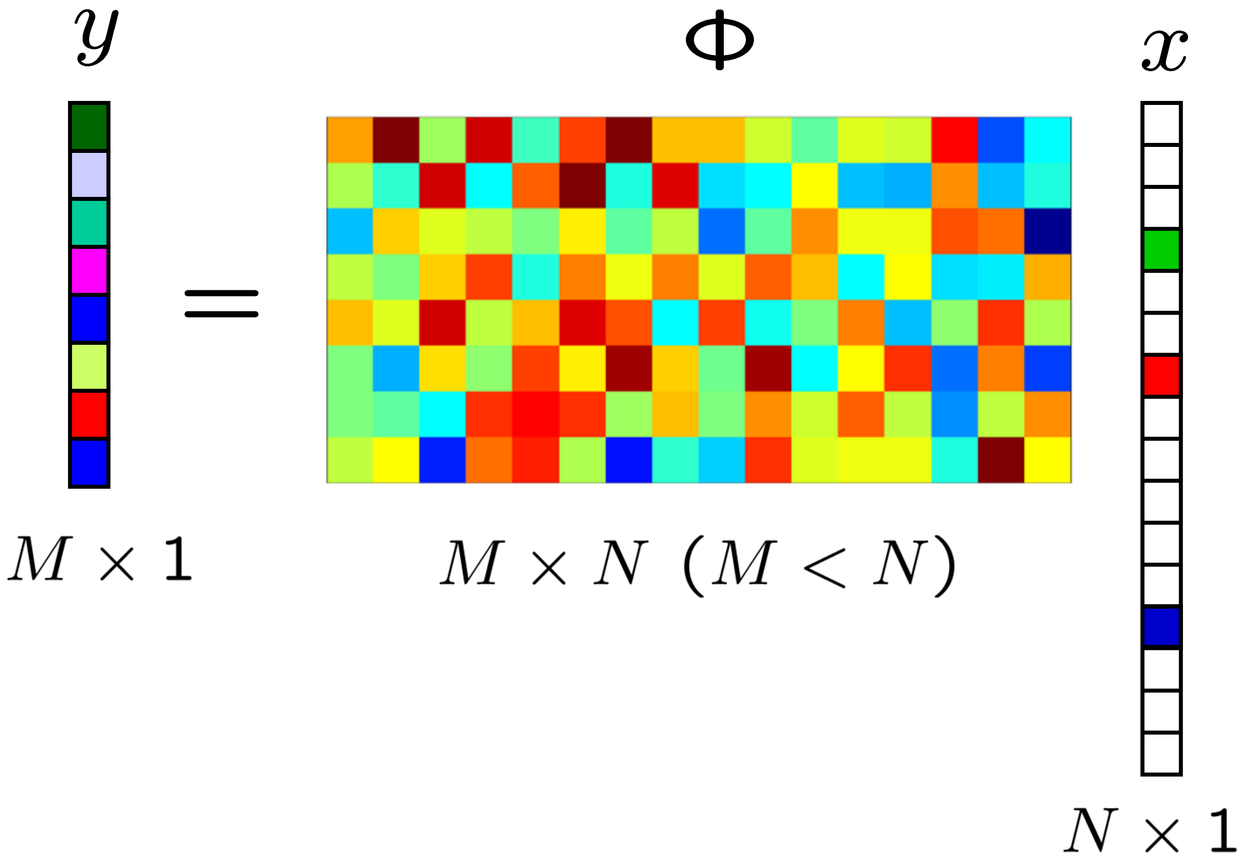
LIONS/Laboratory for Information and Inference Systems

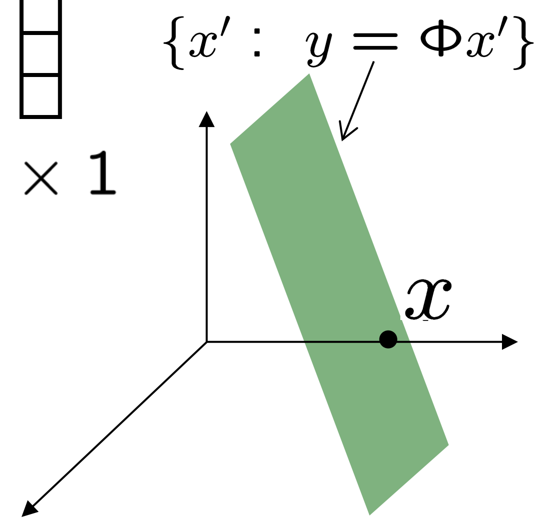
Dimensionality Reduction



Compressive sensing	non-adaptive measurements
Sparse Bayesian learning	dictionary of features
Information theory	coding frame
Theoretical computer science	sketching matrix / expander

Dimensionality Reduction





- **Challenge:** Null space of Φ: $\mathcal{N}(\Phi)$

$$x' = x + v \Rightarrow y, \ \forall v \in \mathcal{N}(\Phi)$$

Approaches

	Deterministic	**Probabilistic**
Prior	sparsity compressibility	$f(x)$
Metric	ℓ_p-norm*	likelihood function

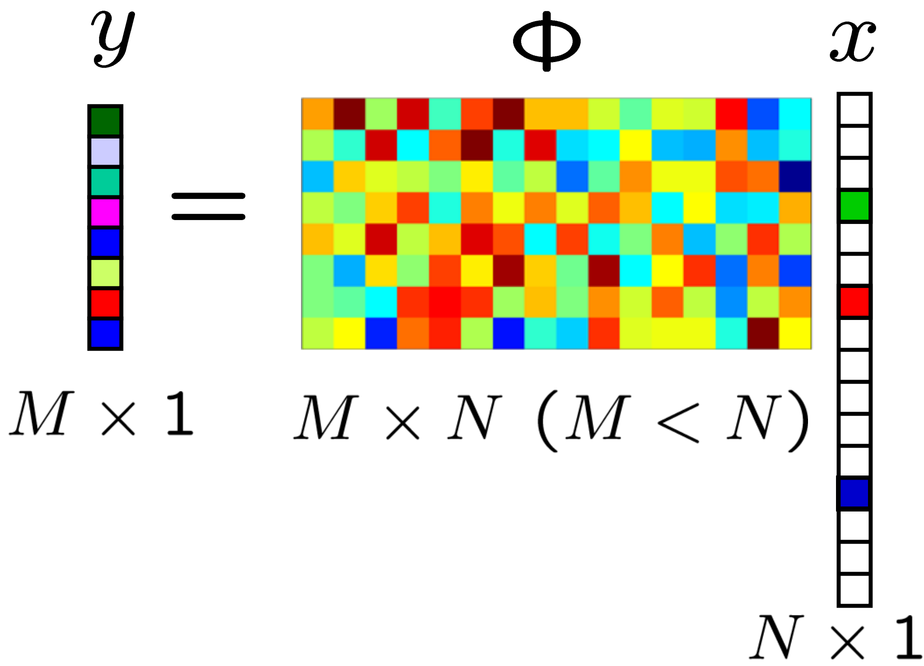
$$* : \|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$$

Deterministic View

My Insights on Compressive Sensing

1. Sparse or compressible x

   not sufficient alone



2. Projection Φ

   information preserving
   (stable embedding / special null space)

3. Decoding algorithms

   tractable

Deterministic Priors

- **Sparse** signal: only K out of N coordinates nonzero
  - model: union of all K-dimensional subspaces aligned w/ coordinate axes

support:

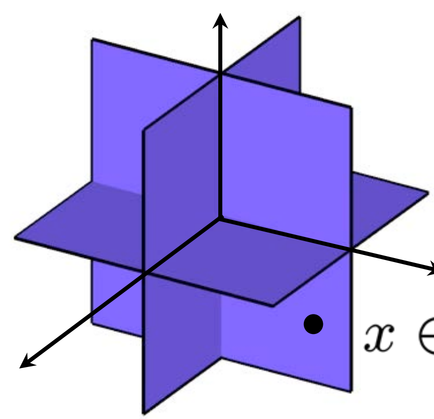
$$\Omega = \{i : x_i \neq 0\}$$

$$|\Omega| = K$$

Example: 2-sparse in 3-dimensions

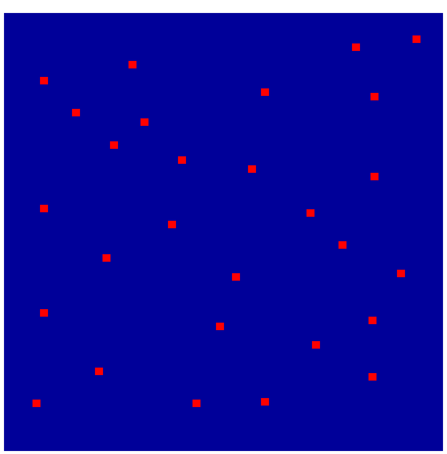
$$K = 2$$

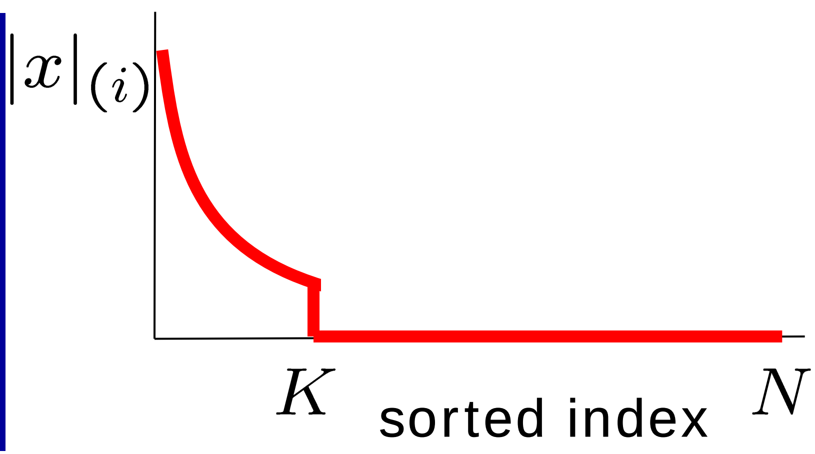
$$\mathbf{R}^3$$

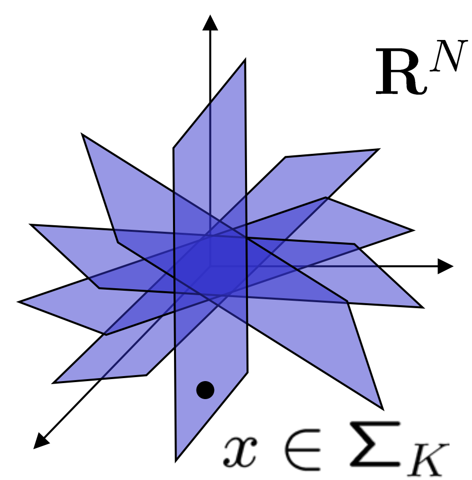


Deterministic Priors

- **Sparse** signal: only K out of N coordinates nonzero
  - model: union of all K-dimensional subspaces aligned w/ coordinate axes

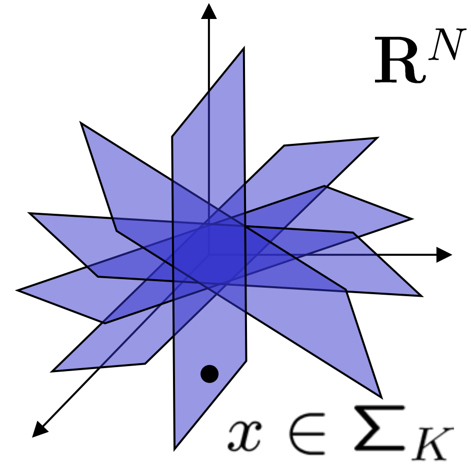






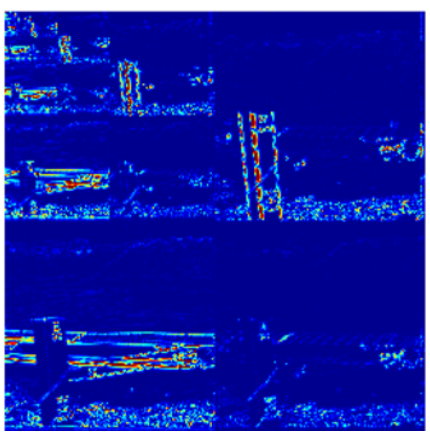
Deterministic Priors

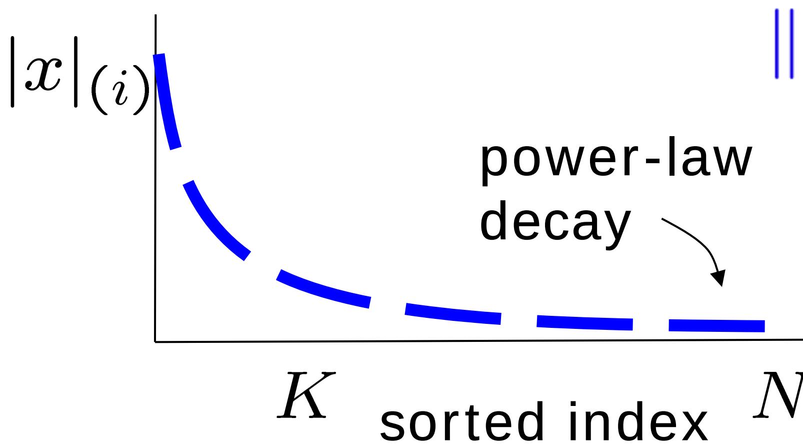
- **Sparse** signal: only K out of N coordinates nonzero
  - model: union of K-dimensional subspaces

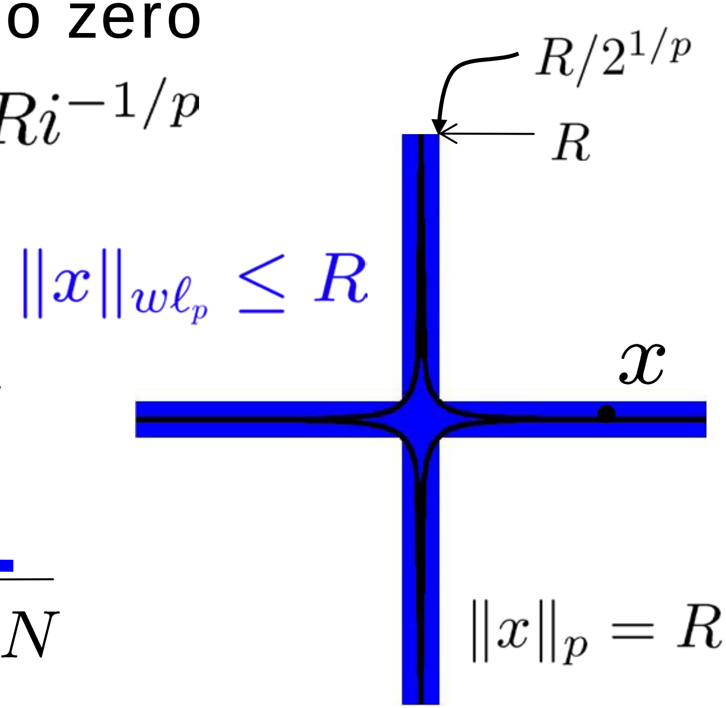


- **Compressible** signal: sorted coordinates decay rapidly to zero
  - Model: weak ℓ_p ball: $|x|_{(i)} \leq R i^{-1/p}$

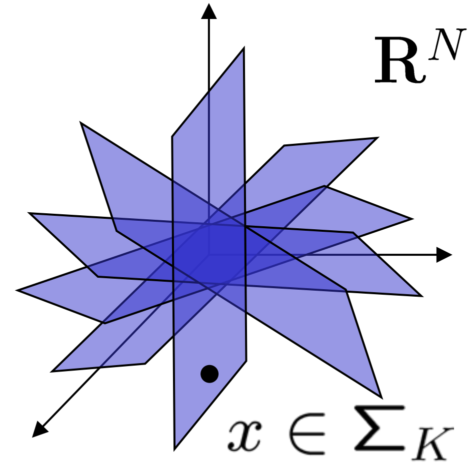
wavelet coefficients:







Deterministic Priors

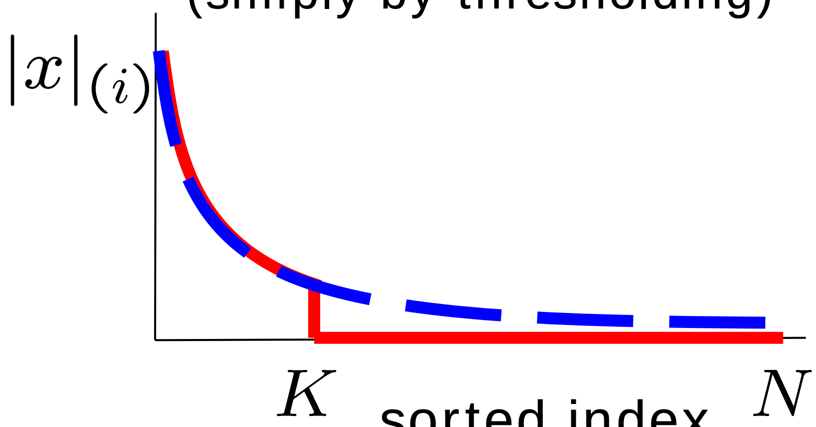


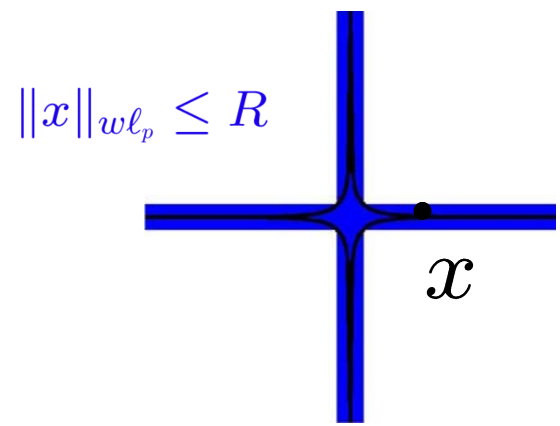
- **Sparse** signal: only K out of N coordinates nonzero
  - model: union of K-dimensional subspaces

- **Compressible** signal: sorted coordinates decay rapidly to zero

$$\|x - x_K\|_r \leq (r/p - 1)^{-1/r} R K^{1/r - 1/p}$$

well-approximated by a K-sparse signal
(simply by thresholding)





Restricted Isometry Property (RIP)

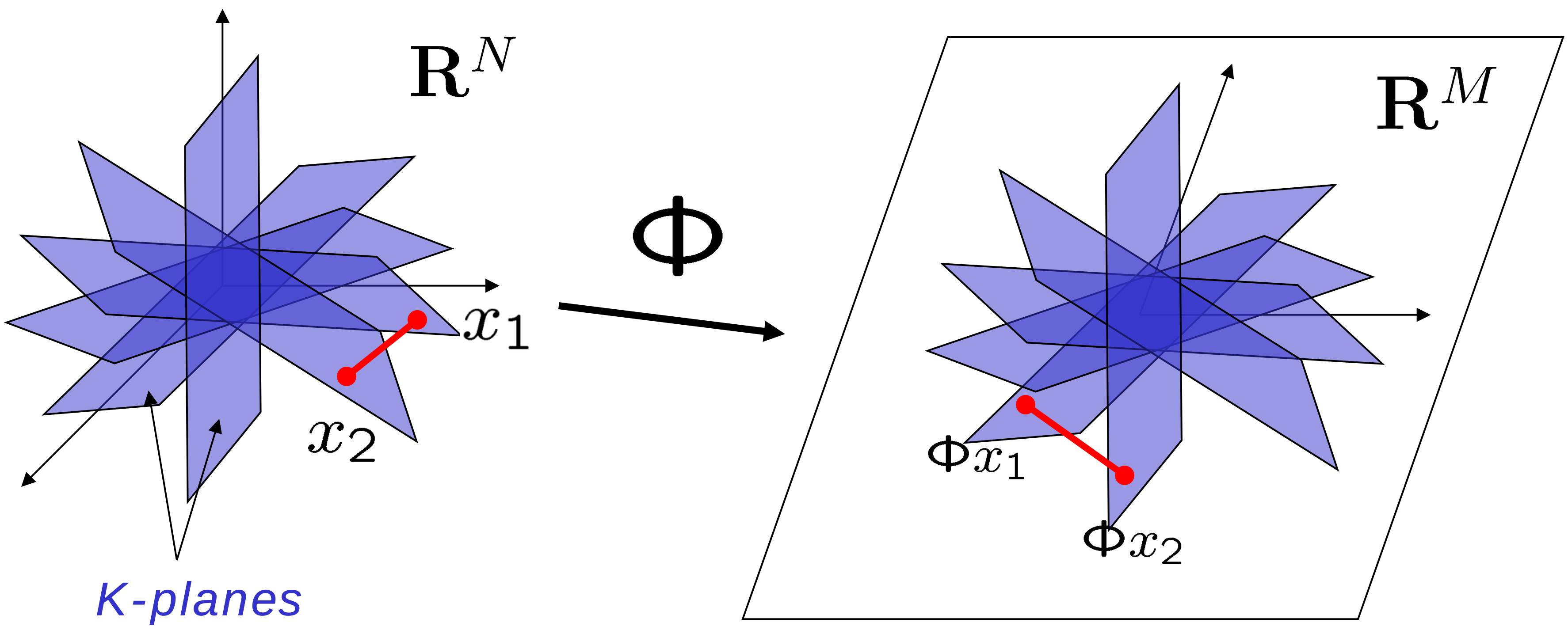
- Preserve the structure of sparse/compressible signals

- RIP of order *2K* implies:
  for all *K*-sparse x_1 and x_2

A random Gaussian matrix has the RIP with high probability if

$$(1-\delta_{2K}) \leq \frac{\|\Phi x_1 - \Phi x_2\|_2^2}{\|x_1 - x_2\|_2^2} \leq (1+\delta_{2K})$$

$$M = O(K \log(N/K))$$



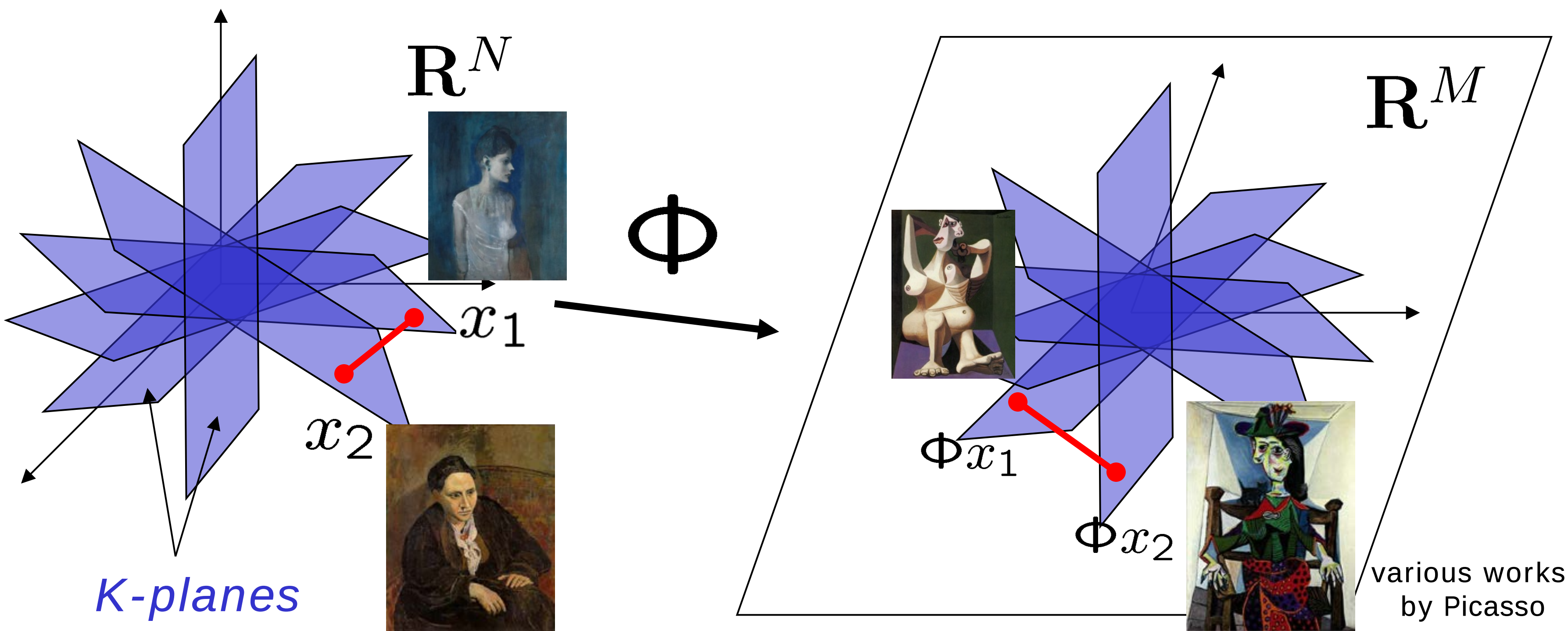
Restricted Isometry Property (RIP)

- Preserve the structure of sparse/compressible signals
- RIP of order *2K* implies:
  for all *K*-sparse x_1 and x_2

$$(1-\delta_{2K}) \leq \frac{\|\Phi x_1 - \Phi x_2\|_2^2}{\|x_1 - x_2\|_2^2} \leq (1+\delta_{2K})$$

A random Gaussian matrix has the RIP with high probability if

$$M = O(K \log(N/K))$$



Robust Null Space Property (RNSP)

- **RNSP in 1-norm (RNSP-1):**

$$\|v_\Omega\|_1 < \eta_K \|v_{\Omega^c}\|_1, \forall v \in \mathcal{N}(\Phi)$$

Ω: support of x

Null space of Φ: $\mathcal{N}(\Phi)$

$\{x' : \ y = \Phi x'\}$

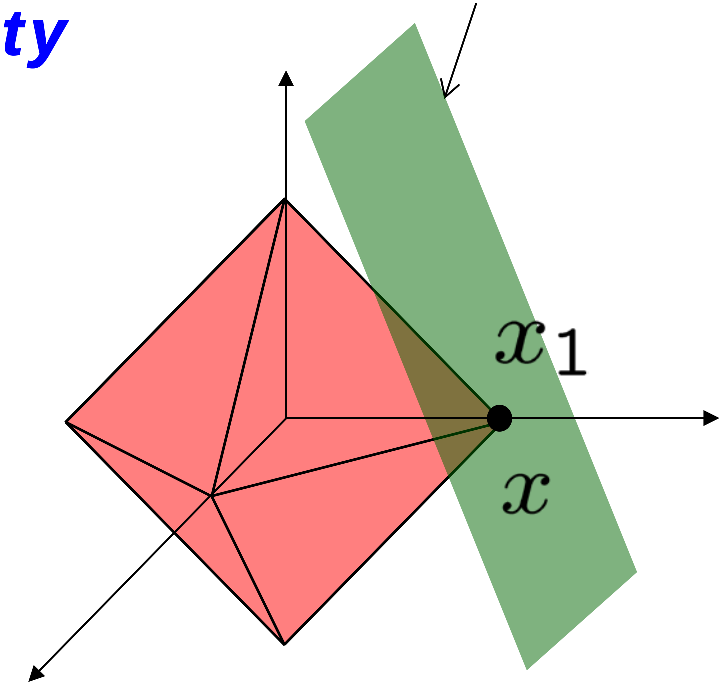
- RNSP-1 < > ***instance optimality***

$$\Delta_1(y) = \underset{x'}{\operatorname{argmin}} \|x'\|_1 \text{ subject to } y = \Phi x'$$

$$\|x - \Delta(\mathbf{\Phi} x)\|_2 \leq 2\frac{1+\eta_K}{1-\eta_K} \cdot \sigma_K(x)_1$$

Best K-term approximation:

$$\sigma_K(x)_q := \inf_{\|u\|_0 \leq K} \|x - u\|_q$$

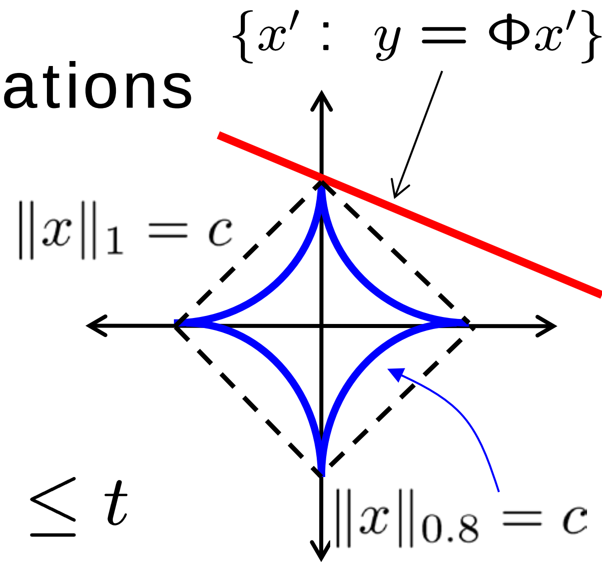


[Cohen, Dahmen, and Devore; Xu and Hassibi; Davies and Gribonval]

Recovery Algorithms

- **Goal:** given $y = \Phi x + n$
  recover x

- $\ell_{q:q\leq 1}$ and convex optimization formulations



  – basis pursuit, Lasso, scalarization ...

$$\widehat{x} = \arg\min \|x\|_q^q \text{ s.t. } y = \Phi x$$

$$\widehat{x} = \arg\min \|y - \Phi x\|_2 \text{ s.t. } \|x\|_q \leq t$$

$$\widehat{x} = \arg\min \|y - \Phi x\|_2^2 + \mu \|x\|_q$$

$M = O(K \log(N/K))$

  – iterative re-weighted $\ell_1 \& \ell_2$ algorithms

- Greedy algorithms: CoSaMP, IHT, SP

Performance of Recovery (q=1)

- **Tractability** *polynomial time*
- **Sparse signals** *instance optimal*
  - noise-free measurements: exact recovery
  - noisy measurements: stable recovery
- **Compressible signals** *instance optimal*
  - recovery as good as *K*-sparse approximation (via RIP)

$$\|x - \widehat{x}\|_2 \leq C_1 \frac{\|x - x_K\|_1}{K^{1/2}} + C_2 \|n\|_2$$

CS recovery error — signal K-term approx error — noise

$$M = O(K \log(N/K))$$

The Probabilistic View

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)}$$

Probabilistic View

- **Goal:** given $y = \Phi x + n$

  recover x

- **Prior:** iid generalized Gaussian distribution (GGD)

  iid: independent and identically distributed

$$f(x) = \text{GGD}(x; q, \lambda) \propto \mathrm{e}^{-(|x|/\lambda)^q}$$

- **Algorithms:** via Bayesian inference arguments

  – maximize prior $\quad \widehat{x} = \arg\min \|x\|_q^q \text{ s.t. } y = \Phi x$

  – prior thresholding $\quad \widehat{x} = \arg\min \|y - \Phi x\|_2 \text{ s.t. } \|x\|_q \le t$

  – maximum a posteriori (MAP) $\quad \widehat{x} = \arg\min \|y - \Phi x\|_2^2 + \mu \|x\|_q^q$

$$(\text{MAP: } n \sim \mathcal{N}(0, \sigma^2) \Rightarrow \mu = 2\sigma^2/\lambda^q)$$

Probabilistic View

- **Goal:** given $y = \Phi x + n$

  recover x

- **Prior:** iid generalized Gaussian distribution (GGD)

$$f(x) = \mathrm{GGD}(x; q, \lambda) \propto \mathrm{e}^{-(|x|/\lambda)^q}$$

- **Algorithms: (q=1 < > deterministic view)**

  $\mathbf{M = O(K \log(N/K))}$

  - maximize prior $\quad \widehat{x} = \arg\min \|x\|_q^q \text{ s.t. } y = \Phi x$
  - prior thresholding $\quad \widehat{x} = \arg\min \|y - \Phi x\|_2 \text{ s.t. } \|x\|_q \leq t$
  - maximum a posteriori (MAP) $\quad \widehat{x} = \arg\min \|y - \Phi x\|_2^2 + \mu \|x\|_q^q$

    $(\text{MAP: } n \sim \mathcal{N}(0, \sigma^2) \Rightarrow \mu = 2\sigma^2/\lambda^q)$

Probabilistic View

- **Goal:** given $y = \Phi x + n$

  recover x

- **Prior:** iid generalized Gaussian distribution (GGD)

$$f(x) = \ \mathrm{GGD}(x; q, \lambda) \propto \mathrm{e}^{-(|x|/\lambda)^q}$$

- **Stable embedding:** **an experiment** by Mike Davies
  - q=1
  - x from N iid samples from GGD (no noise)
  - recover using ℓ_1

Probabilistic View

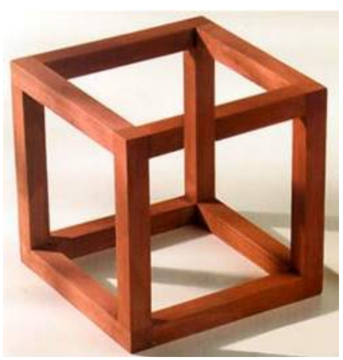
- **Goal:** given $y = \Phi x + n$

  recover x

- **Prior:** iid generalized Gaussian distribution (GGD)

$$f(x) = \mathrm{GGD}(x; q, \lambda) \propto \mathrm{e}^{-(|x|/\lambda)^q}$$

- **Stable Embedding: a paradox**



  - q=1
  - x from N iid samples from GGD (no noise)
  - recover using ℓ_1
  - **need M~0.9 N** (Gaussian Φ) **vs.** $M = O(K \log(N/K))$

Approaches

- Do nothing / Ignore

  be content with
  where we are...

  - generalizes well
  - robust

Compressible Priors*

*You could be a Bayesian if

... your observations are less important than your prior.

[VC; Gribonval, VC, Davies; VC, Gribonval, Davies]

Compressible Priors

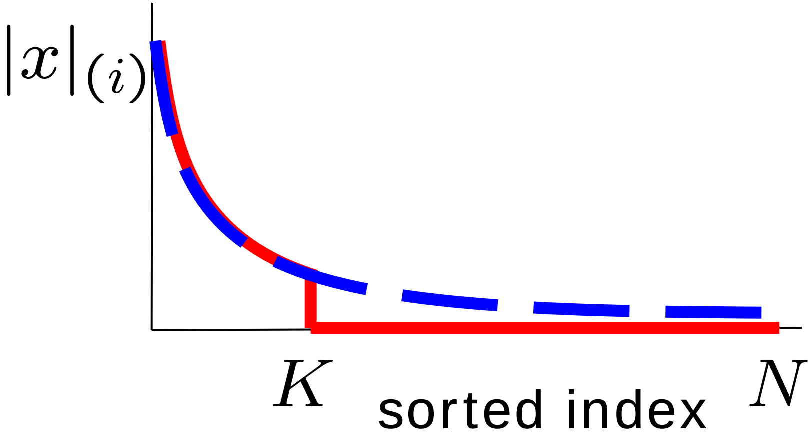
- **Goal:** seek distributions whose iid realizations $x_i \sim p(x)$ can be well-approximated as ***sparse***

Definition:

The PDF $p(x)$ is a *q-compressible prior* with parameters (ϵ, κ), when

$$\lim_{N\to\infty} \bar{\sigma}_{k_N}(x)_q \overset{a.s.}{\leq} \epsilon, \text{(a.s.: almost surely)};$$

for any sequence k_N such that $\lim_{N\to\infty} \inf \frac{k_N}{N} \geq \kappa$, where $\epsilon \ll 1$ and $\kappa \ll 1$.



relative k-term approximation:

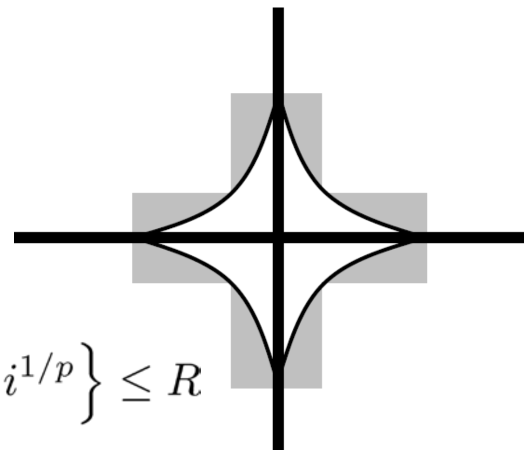
$$\bar{\sigma}_k(x)_q = \frac{\sigma_k(x)_q}{\|x\|_q}$$

$$\sigma_k(x)_q := \inf_{\|u\|_0 \leq k} \|x - u\|_q$$

Compressible Priors

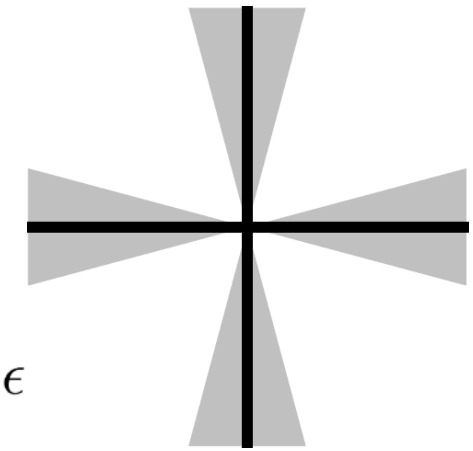
- **Goal:** seek distributions whose iid realizations can be well-approximated as ***sparse***

Classical:



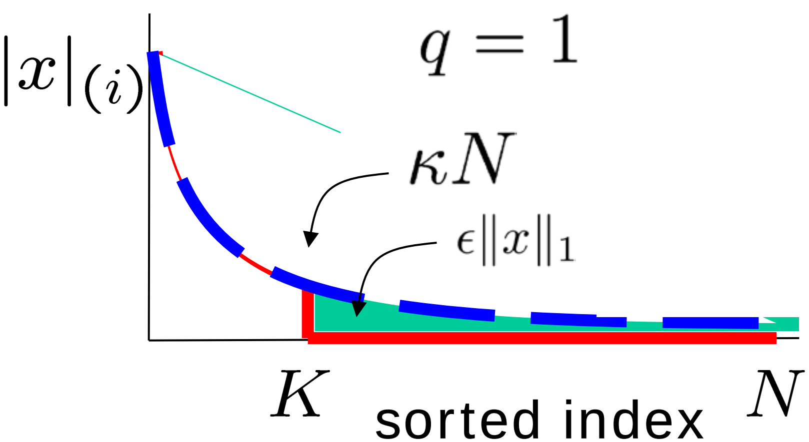
$$\|x\|_{w\ell^p} := \sup_i \left\{ |x|_{(i)} \cdot i^{1/p} \right\} \leq R$$

New:



$$\frac{\sigma_{\kappa N}(x)_q}{\|x\|_q} \leq \epsilon$$

$$q = 1$$



Compressible Priors

- **Goal:** seek distributions whose iid realizations can be well-approximated as ***sparse***
- **Motivations:** deterministic embedding scaffold for the probabilistic view

  analytical proxies for sparse signals
  - learning (e.g., dim. reduced data)
  - algorithms (e.g., structured sparse)

  information theoretic (e.g., coding)
- **Main concept:** ***order statistics***

Key Proposition

Proposition 1. *Suppose* $\mathbf{x}$ *is iid with respect to* $p(x)$. *Denote* $\bar{p}(x) := 0$ *for* $x < 0$, *and* $\bar{p}(x) := p(x) + p(-x)$ *for* $x \geq 0$ *as the PDF of* $|X_n|$, *and* $\bar{F}(t) := \mathbb{P}(|X| \leq t)$ *as its cumulative distribution function. Assume that* $\bar{F}$ *is continuous and strictly increasing on some interval* $[a\ b]$, *with* $\bar{F}(a) = 0$ *and* $\bar{F}(b) = 1$, *where* $0 \leq a < b \leq \infty$. *For any* $0 \leq \kappa \leq 1$, *define the following G-function:*

$$G_q[p](\kappa) := \frac{\int_0^{\bar{F}^{-1}(1-\kappa)} x^q \bar{p}(x)dx}{\int_0^{\infty} x^q \bar{p}(x)dx}. \tag{1}$$

1. ***B****ounded moments: Let* $\mathbb{E}|X|^q < \infty$ *for some* $q \in (0, \infty)$. *Then, given any sequence* k_N *such that* $\lim_{N\to\infty} \frac{k_N}{N} = \kappa \in [0, 1]$, *the following holds almost surely*

$$\lim_{N\to\infty} \bar{\sigma}_k(\mathbf{x})_q^q \overset{a.s.}{=} G_q[p](\kappa). \tag{2}$$

2. ***U****nbounded moments: Let* $\mathbb{E}|X|^q = \infty$ *for some* $q \in (0, \infty)$. *Then, for* $0 < \kappa \leq 1$ *and any sequence* k_N *such that* $\lim_{N\to\infty} \frac{k_N}{N} = \kappa$, *the following holds almost surely*

$$\lim_{N\to\infty} \bar{\sigma}_k(\mathbf{x})_q^q \overset{a.s.}{=} G_q[p](\kappa) = 0. \tag{3}$$

Example 1

- Consider the Laplacian distribution (with scale parameter 1)

$$p_1(x) := \frac{1}{2}\exp\left(-|x|\right)$$

  – The G-function is straightforward to derive

$$G_1[p_1](\kappa) = 1 - \kappa \cdot \Big(1 + \ln 1/\kappa\Big),$$
$$G_2[p_1](\kappa) = 1 - \kappa \cdot \Big(1 + \ln 1/\kappa + \frac{1}{2}(\ln 1/\kappa)^2\Big).$$

- Laplacian distribution < > **NOT** *1 or 2-compressible*

$$\bar{\sigma}_k(\mathbf{x})_1^1 = \frac{\|x - x_K\|_1}{\|x\|_1} \le \epsilon \Rightarrow \kappa = \frac{k_N}{N} \ge (1 - \sqrt{\epsilon})$$

Example 1

- Consider the Laplacian distribution (with scale parameter 1)

$$p_1(x) := \frac{1}{2}\exp\left(-|x|\right)$$

- Laplacian distribution < > **NOT** *1 or 2-compressible*

- **Why does ℓ_1 minimization work for sparse recovery then?**
  - The sparsity enforcing nature of the ℓ_1 cost function
  - The compressible nature of the unknown vector x

Sparse Modeling vs. Sparsity Promotion

- Bayesian interpretation of sparse recovery <> **inconsistent**

four decoding algorithms:

$$\Delta_1(\mathbf{y}) = \underset{\tilde{\mathbf{x}}:\mathbf{y}=\boldsymbol{\Phi}\tilde{\mathbf{x}}}{\operatorname{argmin}} \|\tilde{\mathbf{x}}\|_1,$$

$$\Delta_{\mathrm{LS}}(\mathbf{y}) = \underset{\tilde{\mathbf{x}}:\mathbf{y}=\boldsymbol{\Phi}\tilde{\mathbf{x}}}{\operatorname{argmin}} \|\tilde{\mathbf{x}}\|_2 = \boldsymbol{\Phi}^{+}\mathbf{y},$$

$$\Delta_{\mathrm{oracle}}(\mathbf{y},\Lambda) = \underset{\tilde{\mathbf{x}}:\mathbf{y}=\boldsymbol{\Phi}\tilde{\mathbf{x}},\ \mathrm{support}(\mathbf{x})=\Lambda}{\operatorname{argmin}} \|\tilde{\mathbf{x}}\|_2 = \boldsymbol{\Phi}_{\Lambda}^{+}\mathbf{y},$$

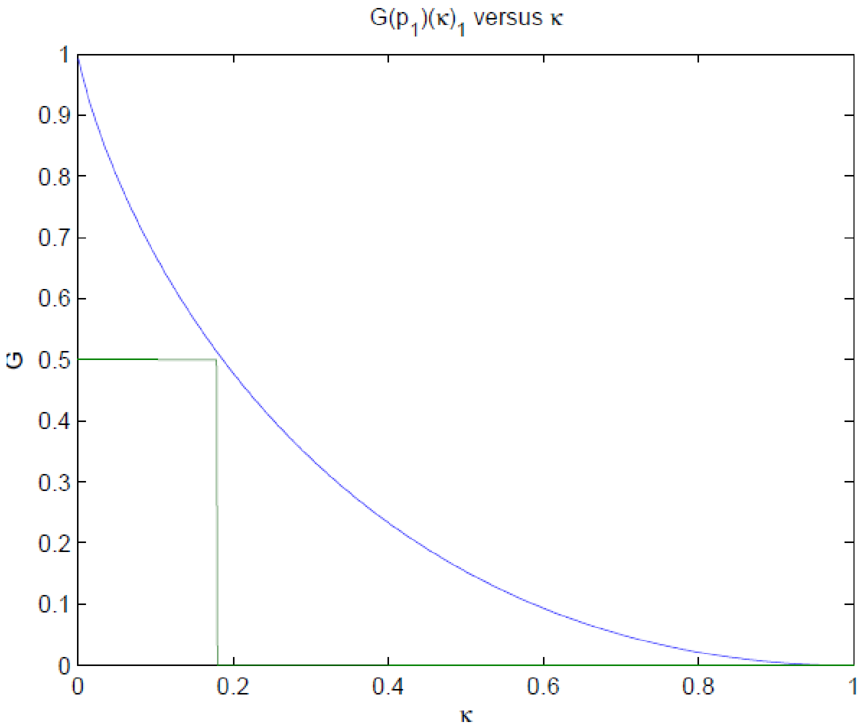
$$\Delta_{\mathrm{trivial}}(\mathbf{y}) = 0,$$

Lemma 1. *Suppose that* $\mathbf{x}$ *is iid with respect to* $p(x)$ *and that* $p(x)$ *satisfies* $G_1[p](\kappa_0) \geq 1/2$*, where* $\kappa_0 \approx 0.18$ *is an absolute constant that depends on the sensing matrix. Then, there is no undersampling ratio* $\delta = m/N$ *for which instance optimality for* Δ_1 *guarantees to outperform the trivial decoder* $\Delta_{trivial}$*.*

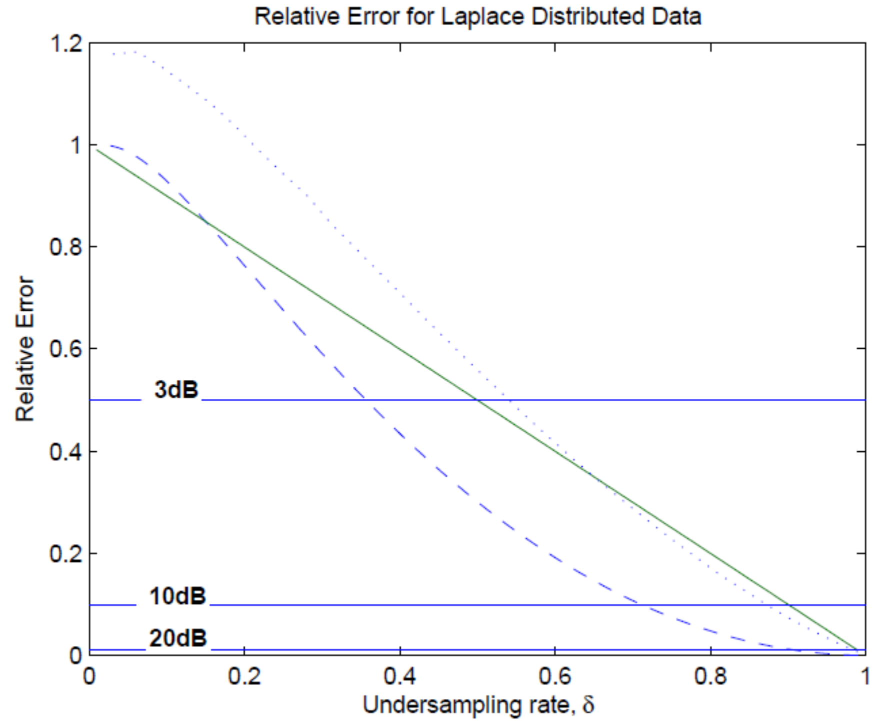
Theorem 1. *Suppose that* $\mathbf{x}$ *is iid with respect to* $p(x)$ *and that* $p(x)$ *has a finite fourth-moment* $\mathbb{E}X^4 < \infty$*. Then there exists a minimum undersampling factor* $\delta_0 = m_0/N$ *such that for any* $\delta < \delta_0$ *and any* k*, the asymptotic performance of oracle* k*-sparse estimation is almost surely worse than that of LS estimation, when* $\mathbf{n} = 0$*.*

($\delta_0 \approx 0.151$: Laplace)

Example 2



G(p1)(κ)1 versus κ
G
κ
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.2
0.4
0.6
0.8



Relative Error for Laplace Distributed Data
Relative Error
Undersampling rate, δ
3dB
10dB
20dB
0
0.2
0.4
0.6
0.8
1
1.2

Approximation Heuristic for Compressible Priors via Order Statistics

- Probabilistic signal model

$$x_i \overset{\text{iid}}{\sim} f(x)$$

$$(\bar{x}_i = |x_i|)$$

< > $\bar{x}_{(1)} \geq \bar{x}_{(2)} \geq \ldots \geq \bar{x}_{(N)}$

order statistics of

$$\bar{f}(\bar{x}) = f(\bar{x}) + f(-\bar{x})$$

Approximation Heuristic for Compressible Priors via Order Statistics

- Probabilistic signal model

  $x_i \overset{\text{iid}}{\sim} f(x)$ <> $\bar{x}_{(1)} \geq \bar{x}_{(2)} \geq \ldots \geq \bar{x}_{(N)}$

  $(\bar{x}_i = |x_i|)$

  order statistics of

  $$\bar{f}(\bar{x}) = f(\bar{x}) + f(-\bar{x})$$

- Deterministic signal model

  $x \in w\ell_p(R)$ <> $\bar{x}_{(i)} \leq R \cdot i^{-1/p}$

Approximation Heuristic for Compressible Priors via Order Statistics

- Probabilistic signal model $x_i \overset{\text{iid}}{\sim} f(x)$ <> $\bar{x}_{(1)} \geq \bar{x}_{(2)} \geq \ldots \geq \bar{x}_{(N)}$

  $(\bar{x}_i = |x_i|)$

  order statistics of

  $$\bar{f}(\bar{x}) = f(\bar{x}) + f(-\bar{x})$$

- Deterministic signal model $x \in w\ell_p(R)$ <> $\bar{x}_{(i)} \leq R \cdot i^{-1/p}$

- Quantile approximation $$\bar{x}_{(i)} \sim \mathcal{N}\left(E[\bar{x}_{(i)}], \frac{\frac{i}{N}\left(1-\frac{i}{N}\right)}{N\left[f\left(E[\bar{x}_{(i)}]\right)\right]^2}\right)$$

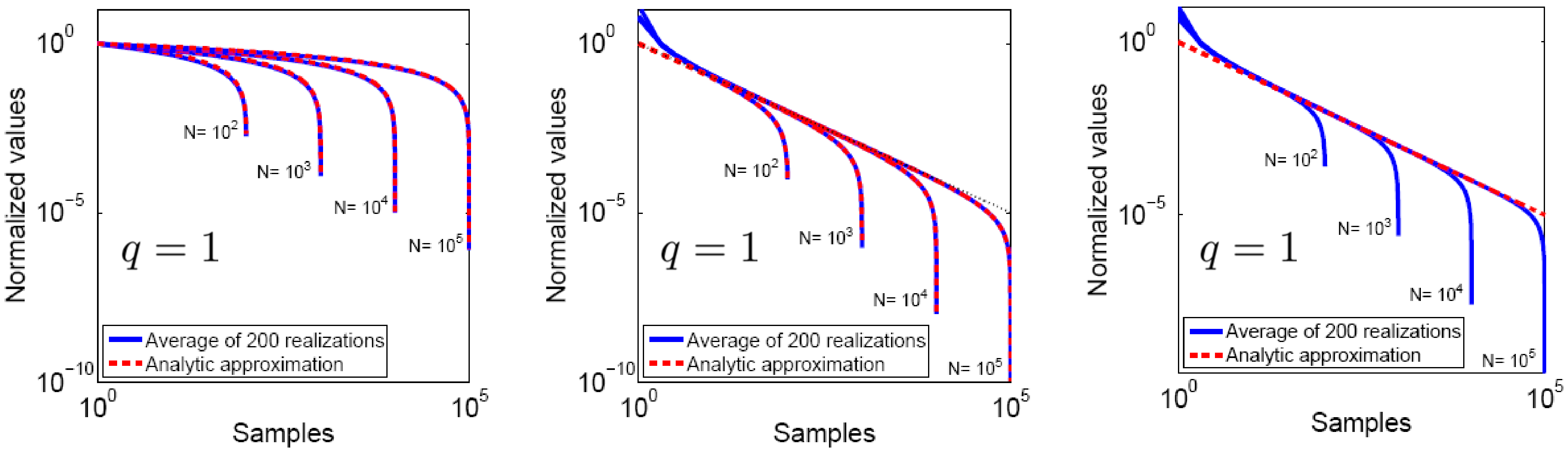
$$R = \bar{F}^{-1}\left(1 - \frac{1}{N}\right),$$
$$p = R\bar{p}(R)N.$$

$$E[\bar{x}_{(i)}] = \bar{F}^{\star}\left(1 - \frac{i}{N+1}\right)$$

$\bar{F}^{\star}(u) = \bar{F}^{-1}(u)$ cdf

Magnitude quantile function (MQF)

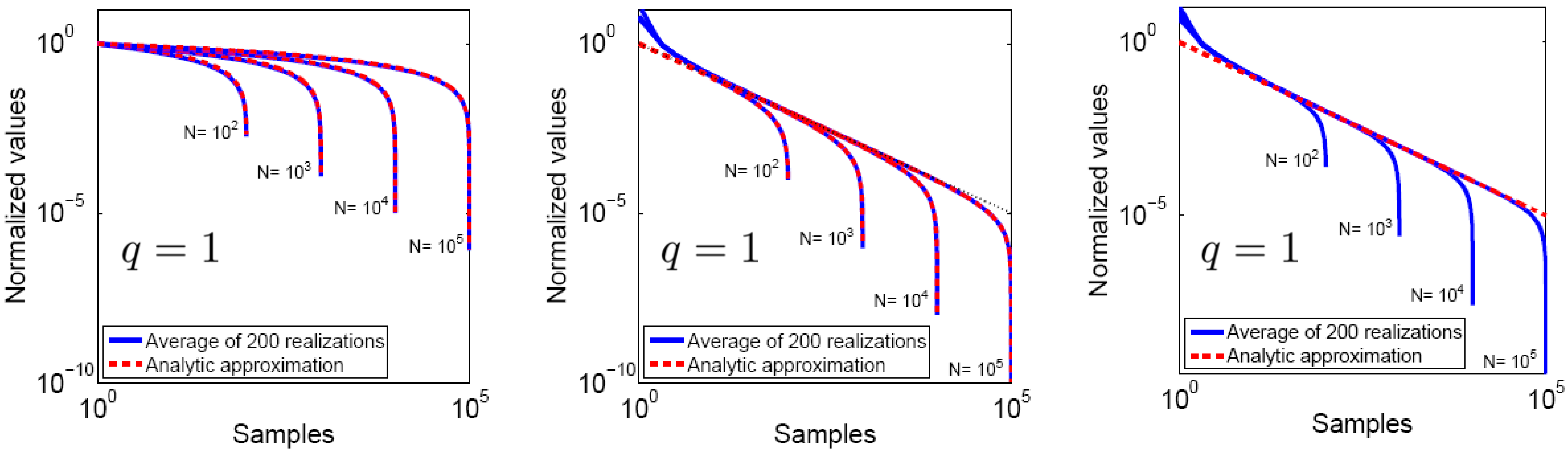
Compressible Priors w/ Examples



(a) GGD (b) GPD (c) Student's t

Distribution	pdf	R	p
Generalized Pareto	$\frac{q}{2\lambda}\left(1+\frac{\lvert x\rvert}{\lambda}\right)^{-(q+1)}$	$\lambda N^{1/q}$	q
Student's t	$\frac{\Gamma((q+1)/2)}{\sqrt{2\pi}\lambda\Gamma(q/2)}\left(1+\frac{x^2}{\lambda^2}\right)^{-(q+1)/2}$	$\left[\frac{2\Gamma((q+1)/2)}{\sqrt{\pi}q\Gamma(q/2)}\right]^{1/q}\lambda N^{1/q}$	q
Fréchet	$(q/\lambda)\,(x/\lambda)^{-(q+1)}\,\mathrm{e}^{-(x/\lambda)^{-q}}$	$\lambda N^{1/q}$	q
Log-Logistic	$\frac{(q/\lambda)(x/\lambda)^{q-1}}{[1+(x/\lambda)^q]^2}$	$\lambda N^{1/q}$	q
Generalized Gaussian	$\frac{q}{2\Gamma(1/q)}\mathrm{e}^{-(\lvert x\rvert/\lambda)^q}$	$\lambda \max\left\{1,\Gamma\left(1+1/q\right)\right\}\log^{1/q}\left(N/q\right)$	$q\log\left(N/q\right)$
Weibull	$(q/\lambda)\,(x/\lambda)^{q-1}\,\mathrm{e}^{-(x/\lambda)^q}$	$\lambda\log^{1/q} N$	$q\log N$
Gamma	$\frac{1}{\lambda\Gamma(q)}\,(x/\lambda)^{q-1}\,\mathrm{e}^{-x/\lambda}$	$\lambda \max\left\{1,\Gamma\left(1+1/q\right)^q\right\}\log\left(qN\right)$	$q\log\left(qN\right)$
Log-Normal	$\frac{q}{\sqrt{2\pi x}}\mathrm{e}^{-(q\log(x/\lambda))^2/2}$	$\lambda\mathrm{e}^{\sqrt{2\log N}/q}$	$\sqrt{2\log N}q$

Compressible Priors w/ Examples



(a) GGD (b) GPD (c) Student's t

Distribution	[redacted]	[redacted]	p
Generalized Pareto	[redacted]	[redacted]	q
Student's t	[redacted]	[redacted]	q
Fréchet	[redacted]	[redacted]	q
Log-Logistic	[redacted]	[redacted]	q
Generalized Gaussian	[redacted]	[redacted]	$q \log (N/q)$
Weibull	[redacted]	[redacted]	$q \log N$
Gamma	[redacted]	[redacted]	$q \log (qN)$
Log-Normal	[redacted]	[redacted]	$\sqrt{2 \log N} q$

Dimensional (in)Dependence

- **Dimensional independence** $p = p(\theta)$ **< >** $M = O(K \log(N/K))$

  < > unbounded moments

example: $K = (p/\epsilon)^{\frac{p}{1-p}} \Rightarrow \|x - x_K\|_1 \leq \epsilon \|x\|_1$

$$\|x - \widehat{x}\|_2 \leq C_1 \frac{\|x - x_K\|_1}{K^{1/2}} + C_2 \|n\|_2$$

CS recovery error

signal K-term approx error

noise

$$M = O(K \log(N/K))$$

Dimensional (in)Dependence

- **Dimensional independence** $p = p(\theta)$ **< >** $M = O(\log N)$

$$K = (p/\epsilon)^{\frac{p}{1-p}} \Rightarrow \|x - x_K\|_1 \leq \epsilon \|x\|_1$$

truly logarithmic embedding

- **Dimensional dependence** $p = p(\theta, N)$ **< >** $M = o(N)$

< > bounded moments

example: iid Laplacian OS: $\bar{x}_{(i)} \approx \lambda \log \frac{N}{i}$

$$K = (1 - \sqrt{\epsilon})N \Rightarrow \|x - x_K\|_1 \leq \epsilon \|x\|_1$$

not so much! / same result can be obtained via the G-function

Why should we care?

- **Natural images**
  - wavelet coefficients

deterministic view vs. **probabilistic view**

Besov spaces

wavelet tresholding

GGD, scale mixtures

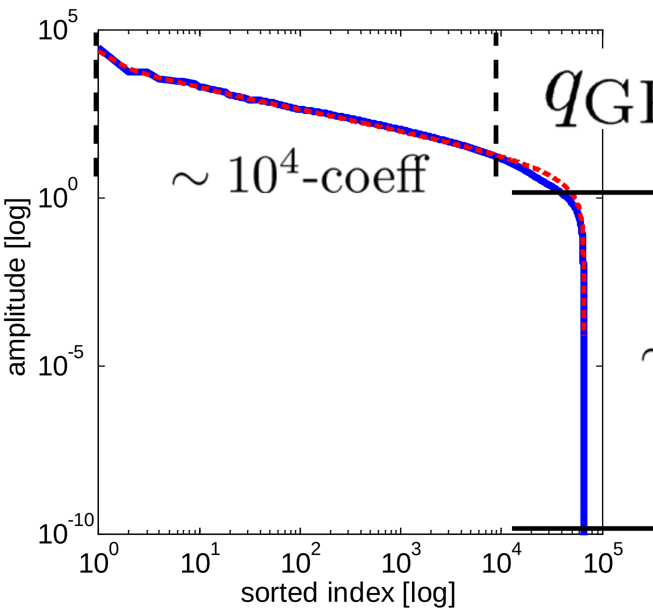
Shannon source coding

Why should we care?

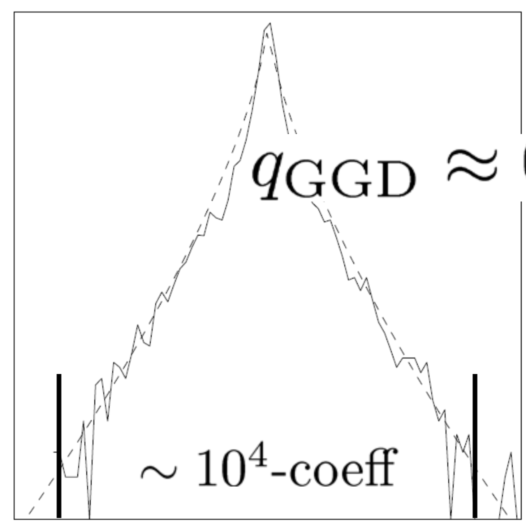
- **Natural images**
  - wavelet coefficients

deterministic view vs. **probabilistic view**

Besov spaces	GGD, scale mixtures
wavelet tresholding	Shannon source coding



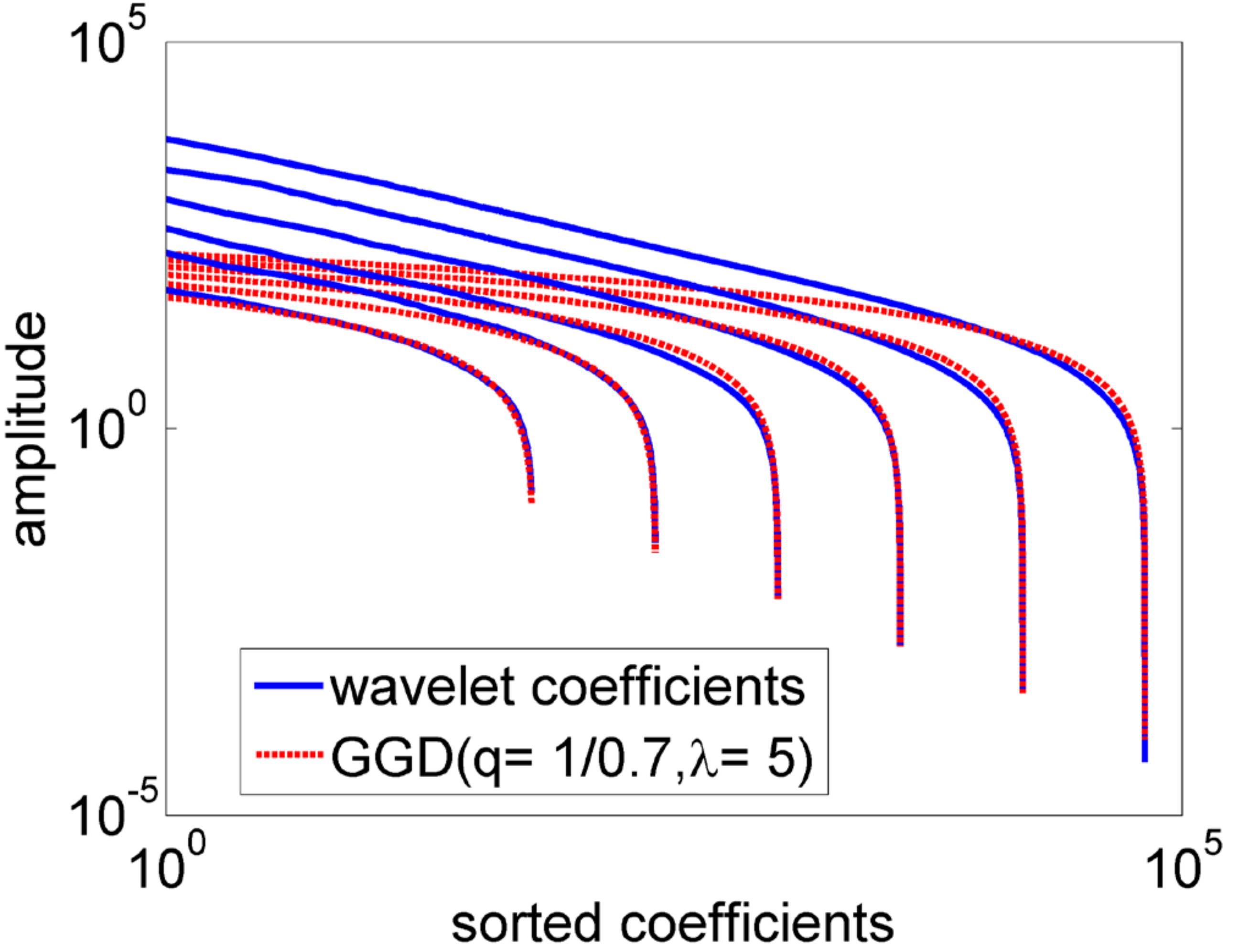
$q_{\text{GPD}} \approx 1.6$



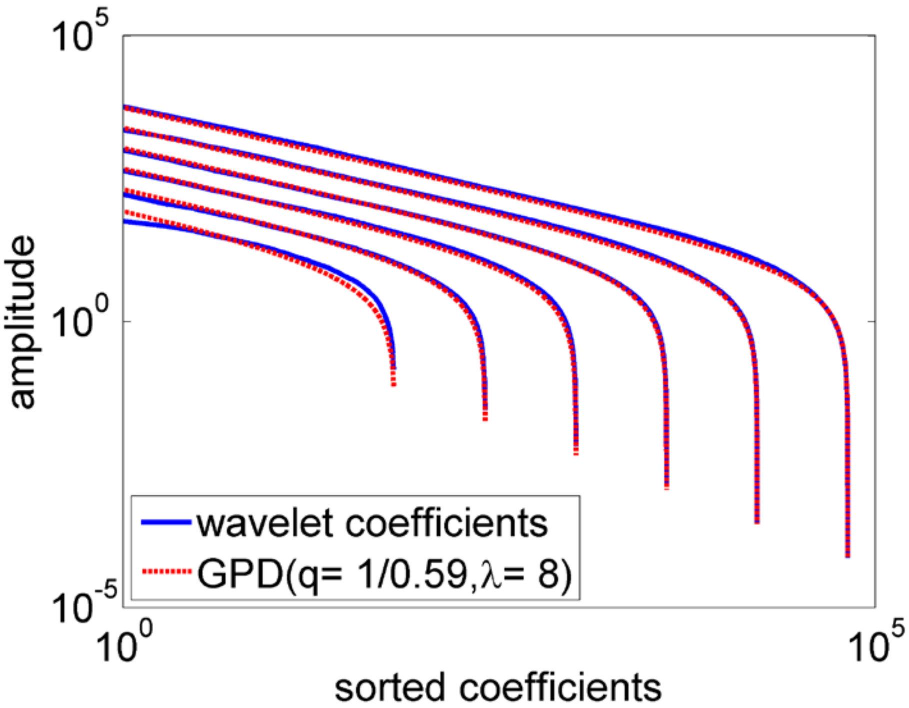
$q_{\text{GGD}} \approx 0.7 - 1$

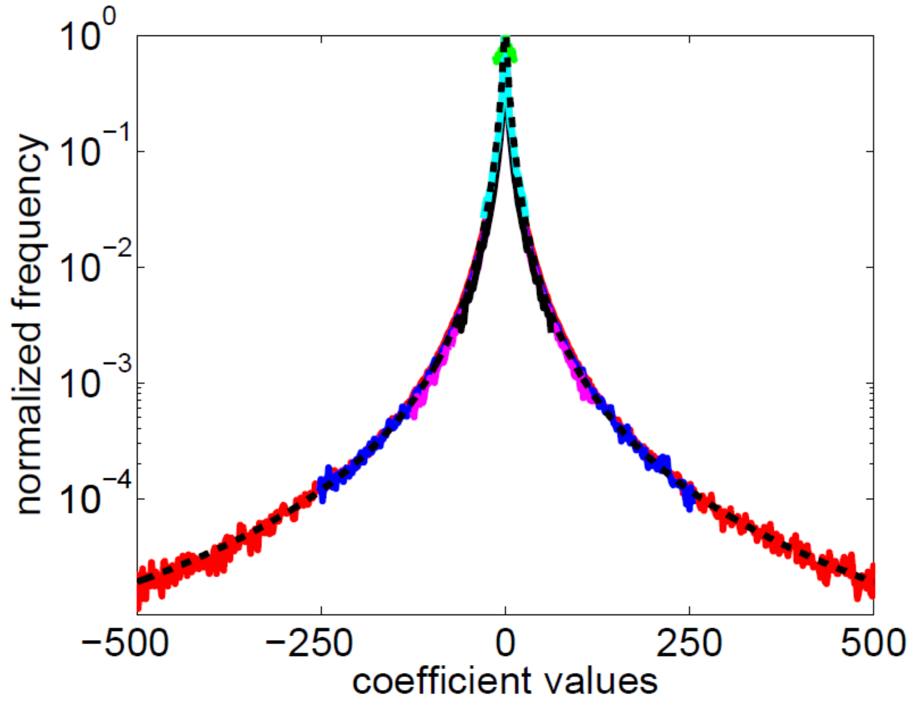
[Choi, Baraniuk; Wainwright, Simoncelli; …]

Berkeley Natural Images Database



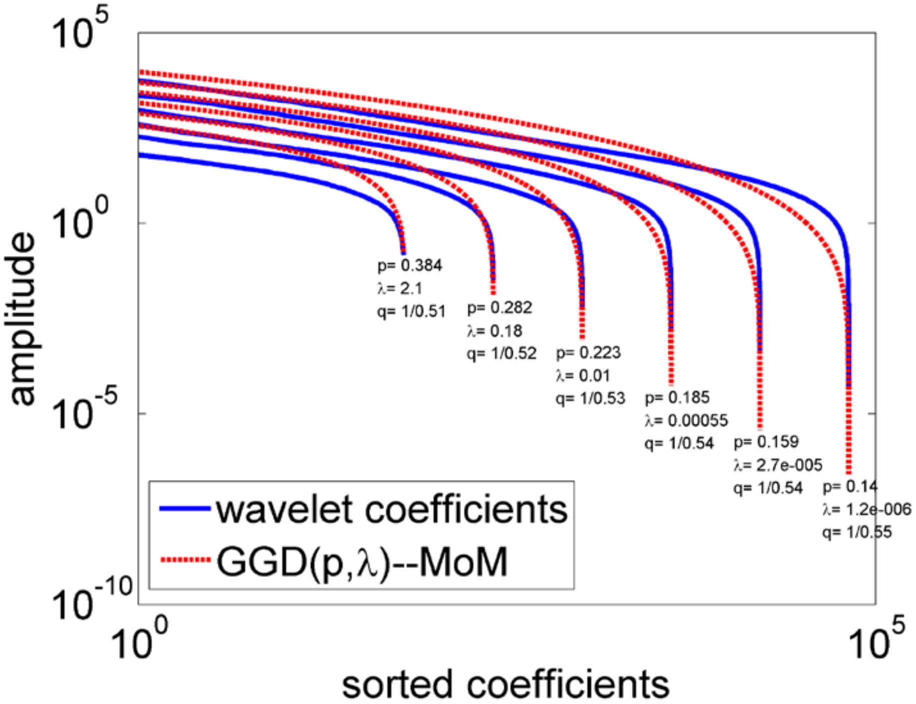
Berkeley Natural Images Database

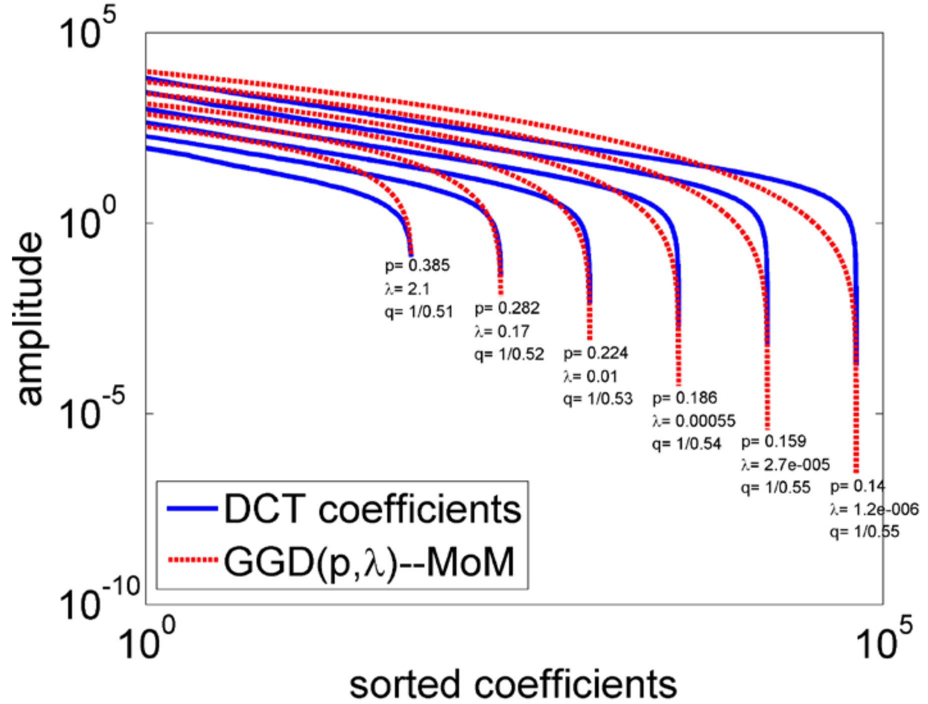




$$\log \mathrm{GPD}(q, \lambda) \doteq -(q+1) \log \left(1 + \frac{|x|}{\lambda}\right) \approx -\frac{|x|}{\lambda/(q+1)}$$

Berkeley Natural Images Database





Learned parameters depend on the dimension

Why should we care?

- **Natural images** (coding / quantization)
  - wavelet coefficients

deterministic view vs. **probabilistic view**

Besov spaces — GGD, scale mixtures

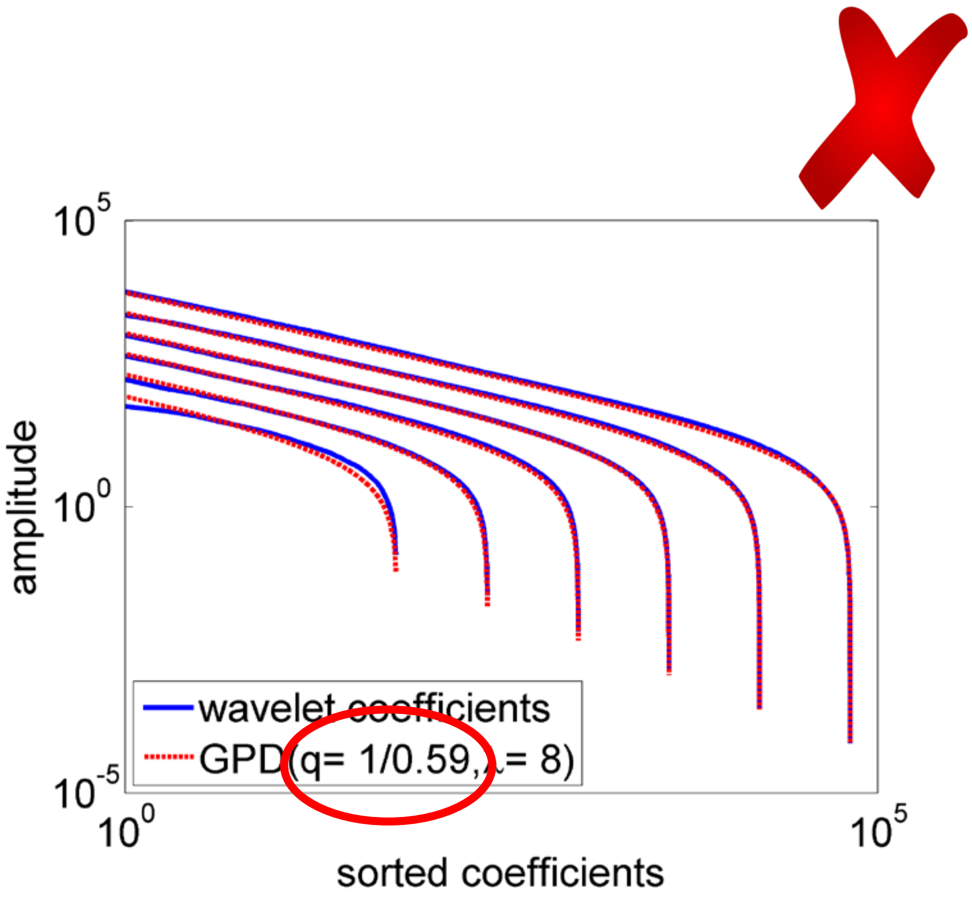
wavelet tresholding — Shannon source coding

(histogram fits, KL divergence)

[bad ideas]

- **Conjecture:** Wavelet coefficients of natural images belong to a dimension independent (non-iid) compressible prior

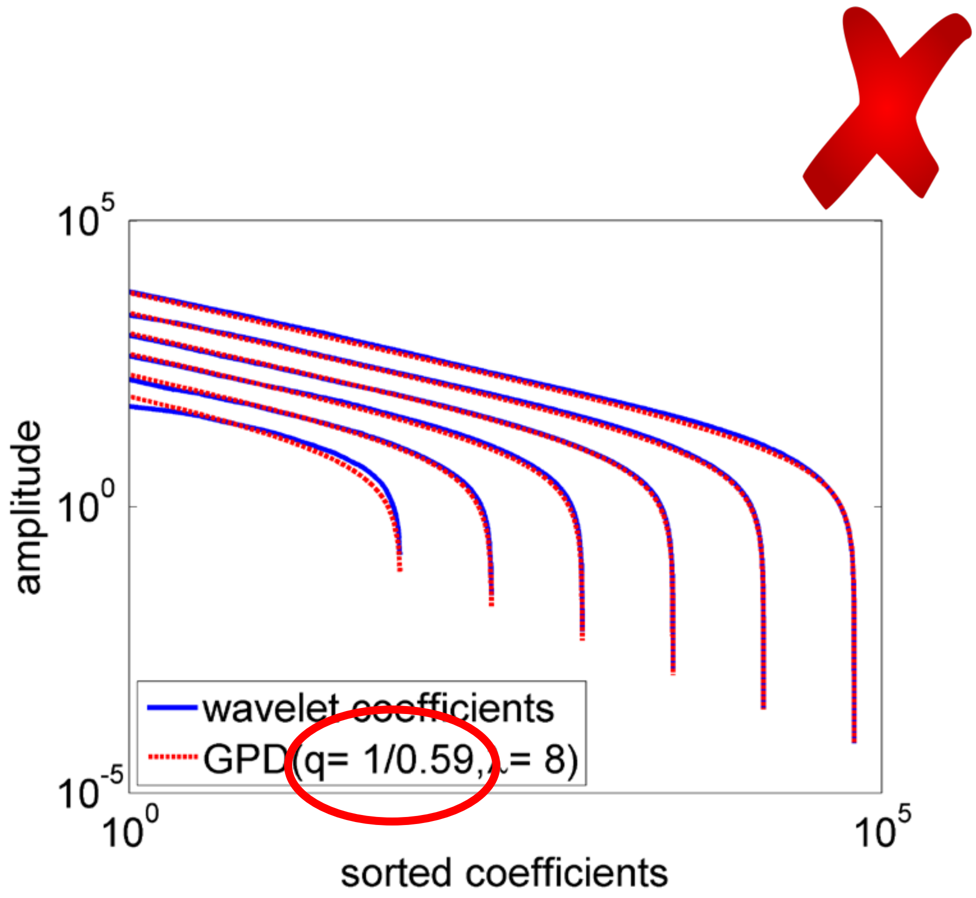
Incompressibility of Natural Images



1-norm instance optimality blows up:

$$\|x - \widehat{x}\|_2 \leq C_1 \frac{\|x - x_K\|_1}{K^{1/2}} + C_2 \|n\|_2$$

Incompressibility of Natural Images



1-norm instance optimality blows up:

$$\|x - \widehat{x}\|_2 \leq C_1 \frac{\|x - x_K\|_1}{K^{1/2}} + C_2 \|n\|_2$$

Is compressive sensing
USELESS
for natural images?

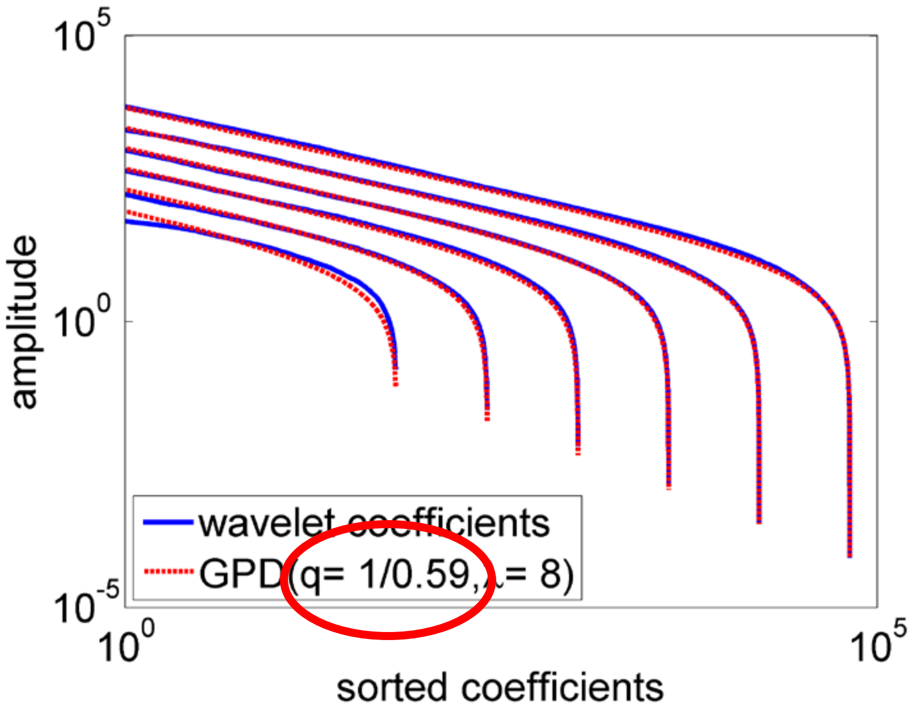
Instance Optimality in Probability to the Rescue

Theorem 2 (Asymptotic performance of the ℓ_1 decoder under infinite second moment). *Let $X_n, n \in \mathbb{N}$ be iid samples from a distribution with PDF $p(x)$ satisfying the hypotheses of Proposition 1. Assume that $\mathbb{E}X^2 = \infty$, and define the coefficient vector $\mathbf{x}_N = (X_1, \ldots, X_N) \in \mathbb{R}^N$. Similarly let $\phi_{i,j}$, $i, j \in \mathbb{N}$ be iid Gaussian variables $\mathcal{N}(0,1)$ and define the $m_N \times N$ Gaussian random matrix $\mathbf{\Phi}_N = \left[\phi_{ij}/\sqrt{m_N}\right]_{1\le i\le m_N, 1\le j\le N}$.*

Consider a sequence of integers m_N such that $\lim_{N\to\infty} m_N/N = \delta$ then

$$\frac{\|\Delta_1(\mathbf{\Phi}_N \mathbf{x}_N) - \mathbf{x}_N\|_2}{\|\mathbf{x}_N\|_2} \overset{a.s.}{\to} 0$$

Incompressibility of Natural Images

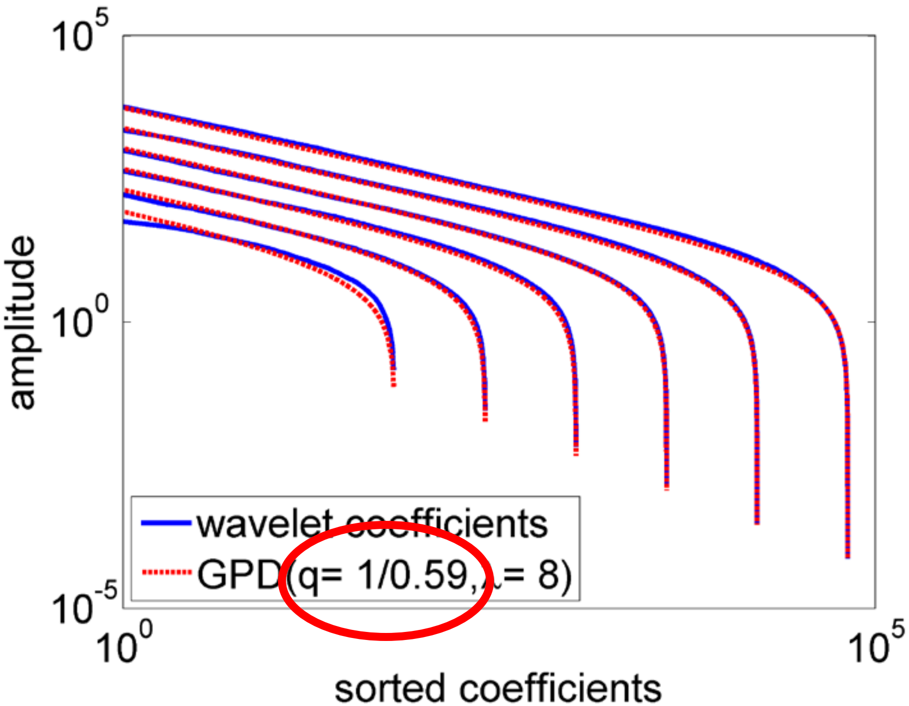


Is compressive sensing USELESS for natural images?

Not according to Theorem 2!!!

For large *N*, 1-norm minimization is still near-optimal.

Incompressibility of Natural Images

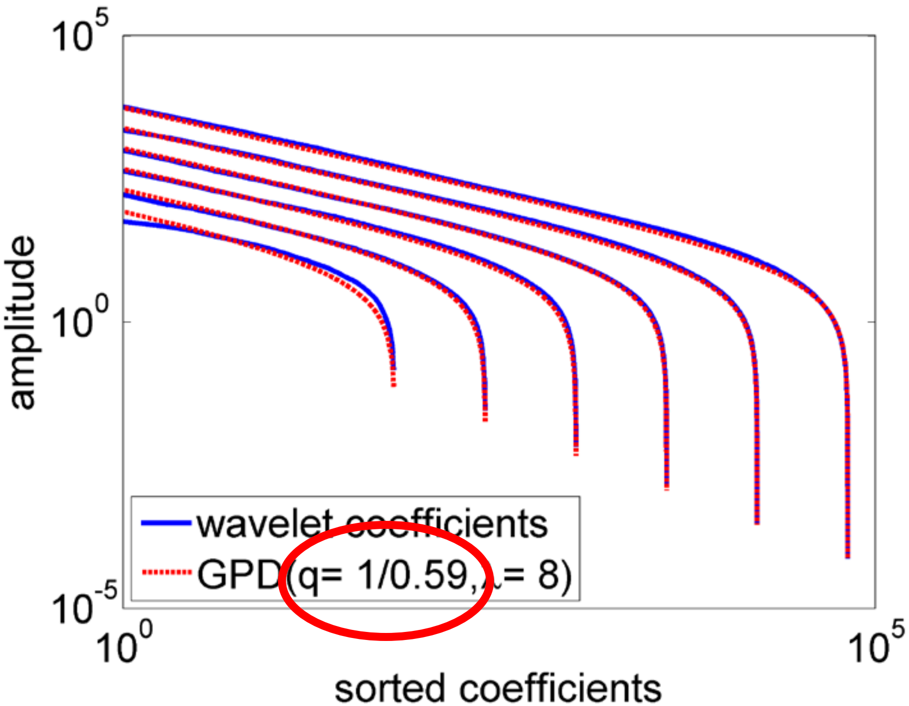


Is compressive sensing
USELESS
for natural images?

Not according to Theorem 2!!!

But, are we NOT missing something practical?

Incompressibility of Natural Images



But, are we NOT missing something practical?

Natural images have finite energy since we have finite dynamic range.

While the resolution of the images are currently ever-increasing, their dynamic range is not.

In this setting, compressive sensing using naïve sparsity will not be useful.

Other Bayesian Interpretations

- Multivariate Lomax dist. (non-iid, compressible w/ r=1)

$$f(x_1, \ldots, x_N) \propto \frac{1}{\left(1+\sum_i \lambda_i^{-1}|x_i|\right)^{q+N}}$$

(has GPD(x_i;q,λ_i) marginals)

$\lambda_i = \lambda$

  - maximize prior

$$\widehat{x} = \arg\min \|x\|_1 \text{ s.t. } y = \Phi x$$

  - prior thresholding

$$\widehat{x} = \arg\min \|y - \Phi x\|_2 \text{ s.t. } \|x\|_1 \leq t$$

  - maximum a posteriori (MAP)

$$\widehat{x}^{\{k\}} = \arg\min \|y - \Phi x\|_2^2 + \mu^{\{k\}} \|x\|_1$$

$$(n \sim \mathcal{N}(0, \sigma^2) \Rightarrow \mu^{\{k\}} = 2\sigma^2(q+N)/(\lambda + \|\widehat{x}^{\{k-1\}}\|_1))$$

fixed point continuation

- Interactions of Gamma and GGD

$$f(x) \propto \frac{1}{(1+|x|^r/\lambda^r)^{\frac{q+1}{r}}}$$

  - iterative re-weighted ℓ_r algorithms

Summary of Results

Table 1: Simple Rule of Thumbs for IID Compressibility and Linear Regression

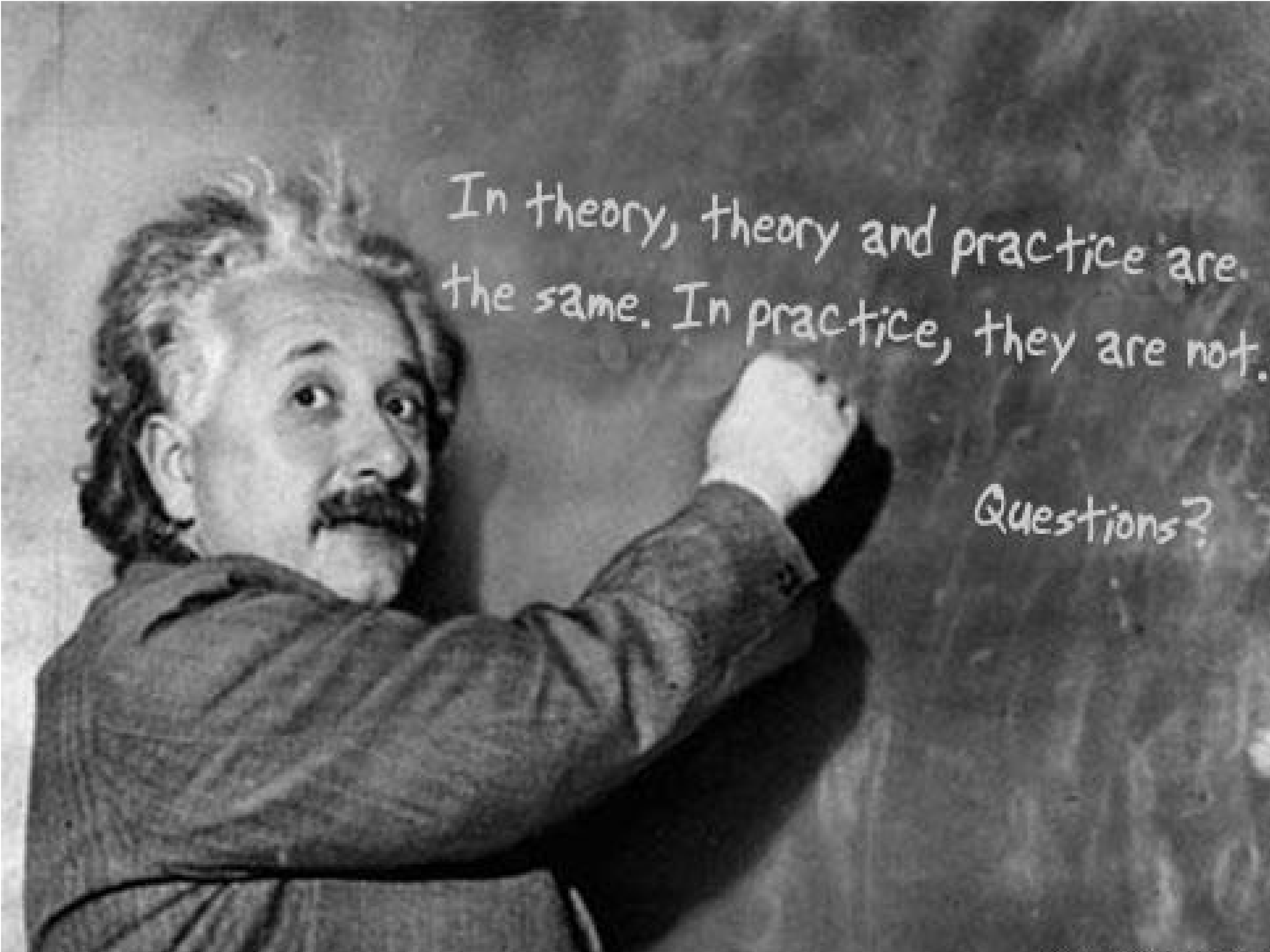
Moment property	$\mathbb{E}x^2 = \infty$	$\mathbb{E}x^2 < \infty$ and $\mathbb{E}x^4 = \infty$	$\mathbb{E}x^4 < \infty$
General result	Δ_1 performs ideally for any δ	N/A depends on finer properties of $p(x)$	Δ_{LS} outperforms Δ_{oracle} for small $\delta < \delta_0$
Examples		Example: $p_0(x) := 2\|x\|/(x^2+1)^3$ Δ_{oracle} performs just as Δ_{LS}	Example: $p_{\tau,\lambda}(x) \propto \exp\left(-\|x/\lambda\|^\tau\right)$ $0 < \tau < \infty$ Generalized Gaussian
	Example: $p(x) \propto \left(1 + \|x/\lambda\|^\tau\right)^{-(q+1)/\tau}$ Generalized Pareto ($\tau = 1$) / Student's t ($\tau = 2$)		
	Case $0 < q \leq 2$	Case $2 < q < 4$ Δ_{oracle} outperforms Δ_{LS} for small $\delta < \delta_0$	Case $q > 4$

$$\delta = M/N$$

Conclusions

- Compressible priors < > probabilistic models for compressible signals (deterministic view)
- q-parameter < > (un)bounded moments
  - independent of N — truly logarithmic embedding with tractable recovery

    dimension agnostic learning
  - not independent of N — many restrictions (embedding, recovery, learning)
- Natural images < > CS is not a good idea w/ naïve sparsity
- *Why would compressible priors be more useful vs.* ℓ_1*?*
  - *Ability to determine the goodness or confidence of estimates*


In theory, theory and practice are the same. In practice, they are not.
Questions?