Sampling rates for ℓ^1-synthesis

"Combien de mesures sous-gaussiennes doit-on faire pour reconstruire un objet parcimonieux dans un dictionnaire redondant ?"

Claire Boyer

November 6th, 2020

Joint work with

Maximilian März
(TU Berlin)

Jonas Kahn
(IMT Toulouse)

Pierre Weiss
(IMT Toulouse)

Outline

Summary

Setting

Linear Noisy Measurements

- Signal: $\boldsymbol{x}_0 \in \mathbb{R}^n$
- Measurements: $\boldsymbol{y} \in \mathbb{R}^m$ of $\boldsymbol{x}_0$ via the linear acquisition model

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{e}, \tag{1}$$

where
$\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a **Gaussian measurement matrix**
$\boldsymbol{e} \in \mathbb{R}^m$ models **measurement noise** with $\|\boldsymbol{e}\|_2 \leq \eta$ for some $\eta \geq 0$

Linear Noisy Measurements

- Signal: $\boldsymbol{x}_0 \in \mathbb{R}^n$
- Measurements: $\boldsymbol{y} \in \mathbb{R}^m$ of $\boldsymbol{x}_0$ via the linear acquisition model

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{e}, \tag{1}$$

where
$\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a **Gaussian measurement matrix**
$\boldsymbol{e} \in \mathbb{R}^m$ models **measurement noise** with $\|\boldsymbol{e}\|_2 \leq \eta$ for some $\eta \geq 0$

Gaussian assumption

- classical benchmark setup in CS
- It allows us to determine the sampling rate of a convex program (i.e., the number of required measurements for successful recovery) by calculating the so-called Gaussian mean width

The signal structure

As for the signal $\boldsymbol{x}_0$

- sparsity hardly satisfied in any real-world application
- but sparse representations using specific transforms
  $\rightsquigarrow$ Gabor dictionaries, wavelet systems or data-adaptive representations

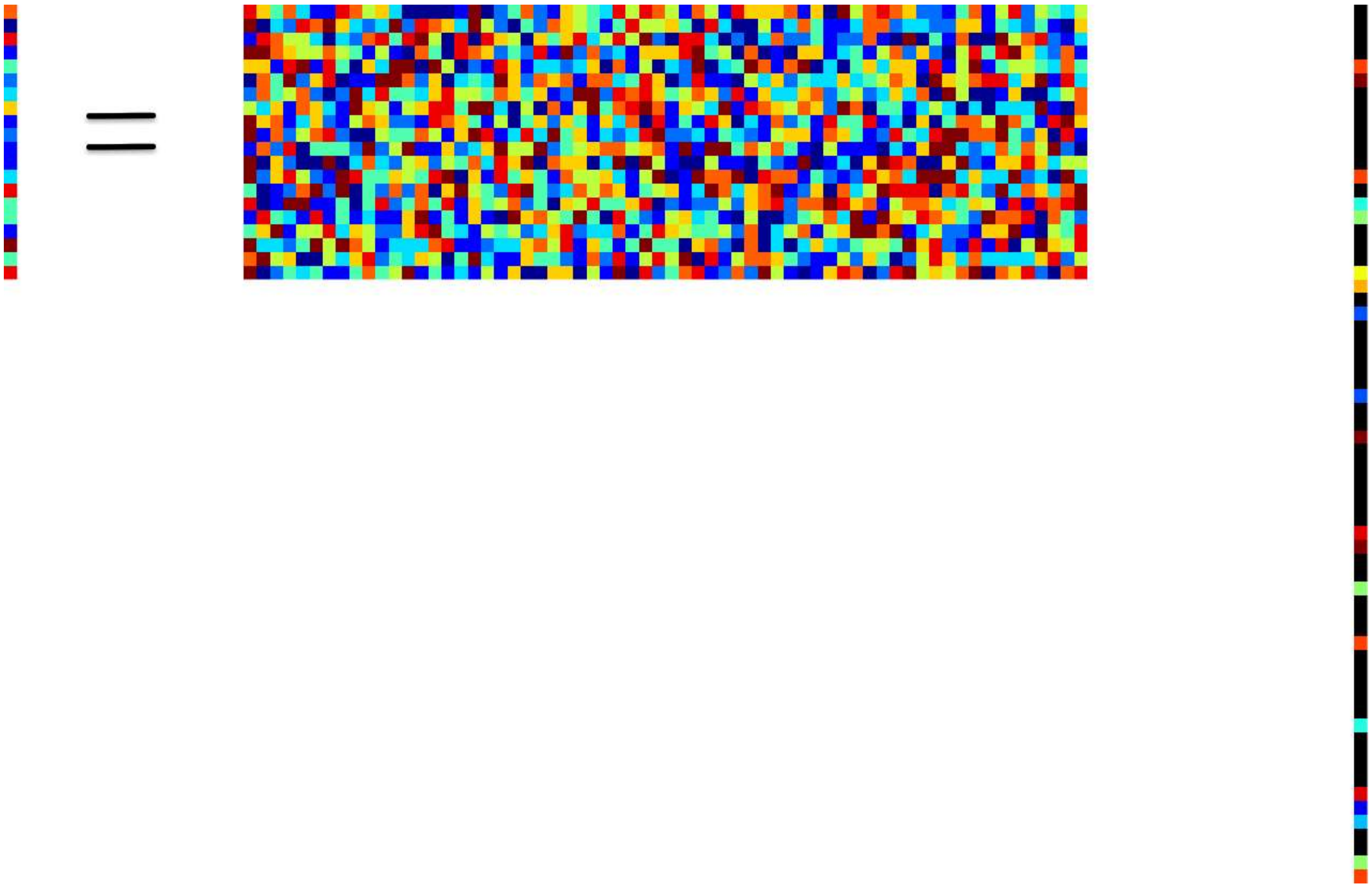
Synthesis formulation

There exists a matrix $\boldsymbol{D} \in \mathbb{R}^{n \times d}$ and a low-complexity representation $\boldsymbol{z}_0 \in \mathbb{R}^d$ such that $\boldsymbol{x}_0$ can be "synthesized" as

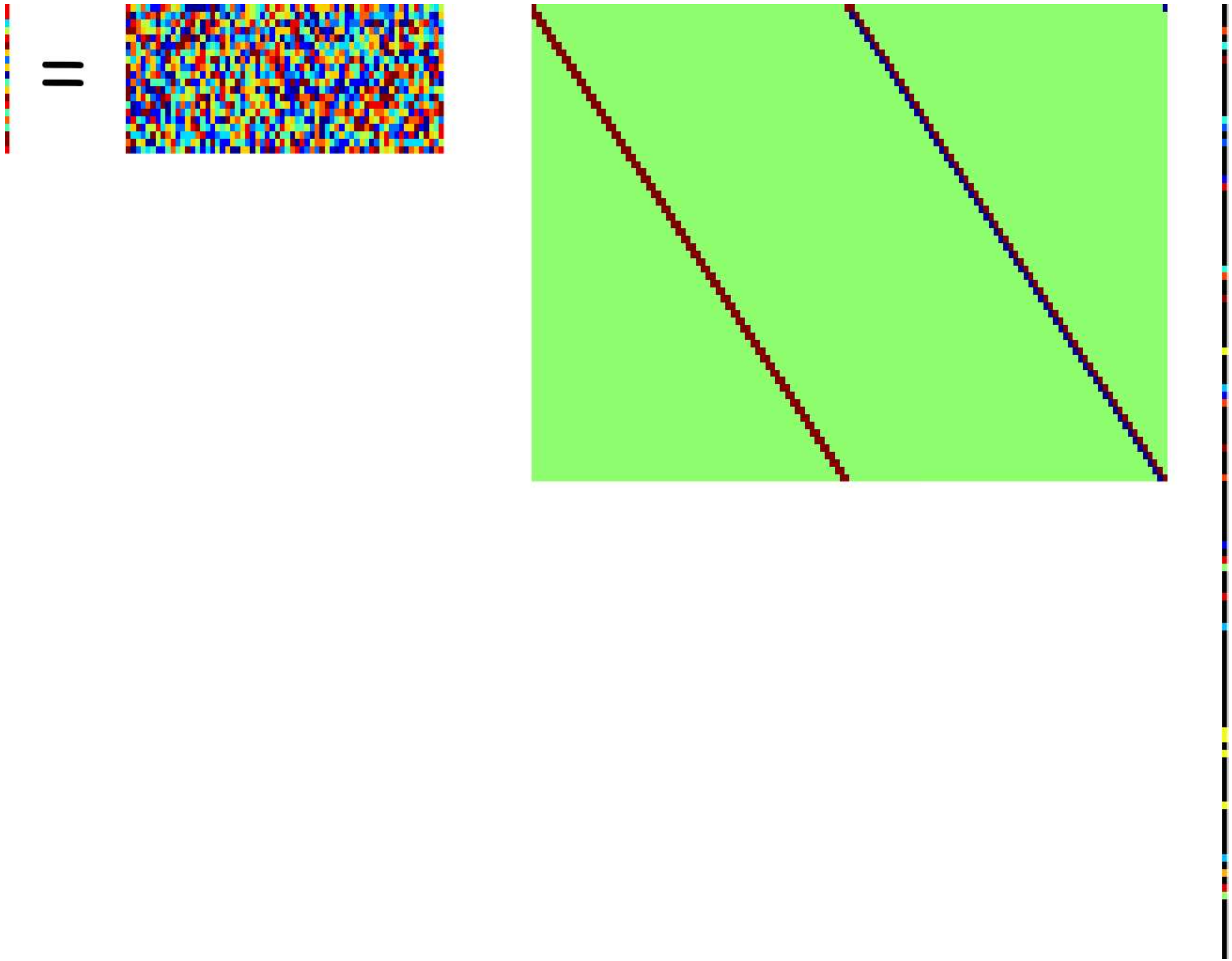
$$\boldsymbol{x}_0 = \boldsymbol{D} \cdot \boldsymbol{z}_0.$$

- $\boldsymbol{D} = [\boldsymbol{d}_1, \dots, \boldsymbol{d}_d]$ is the dictionary
- its columns are the dictionary atoms.

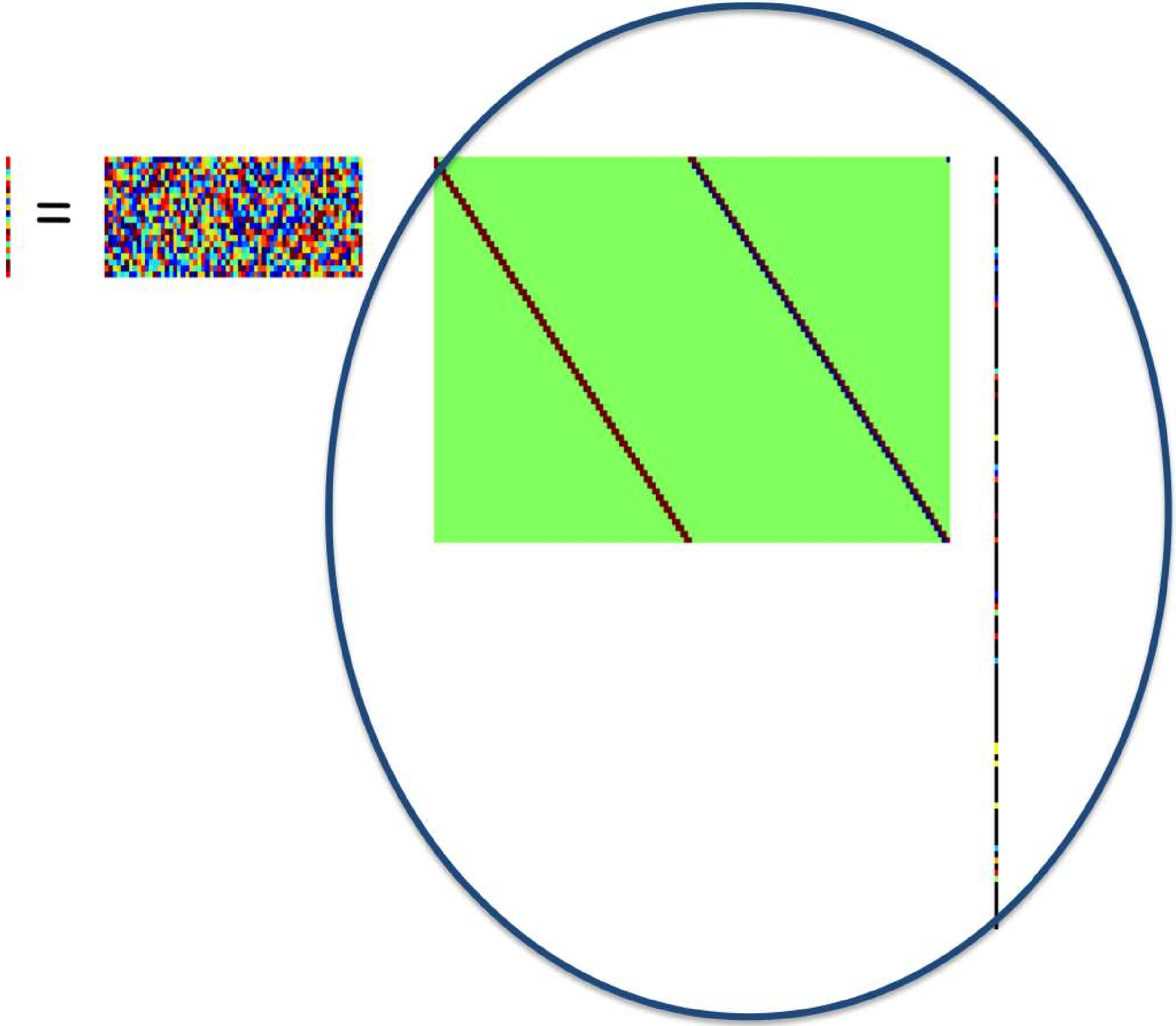
Visually



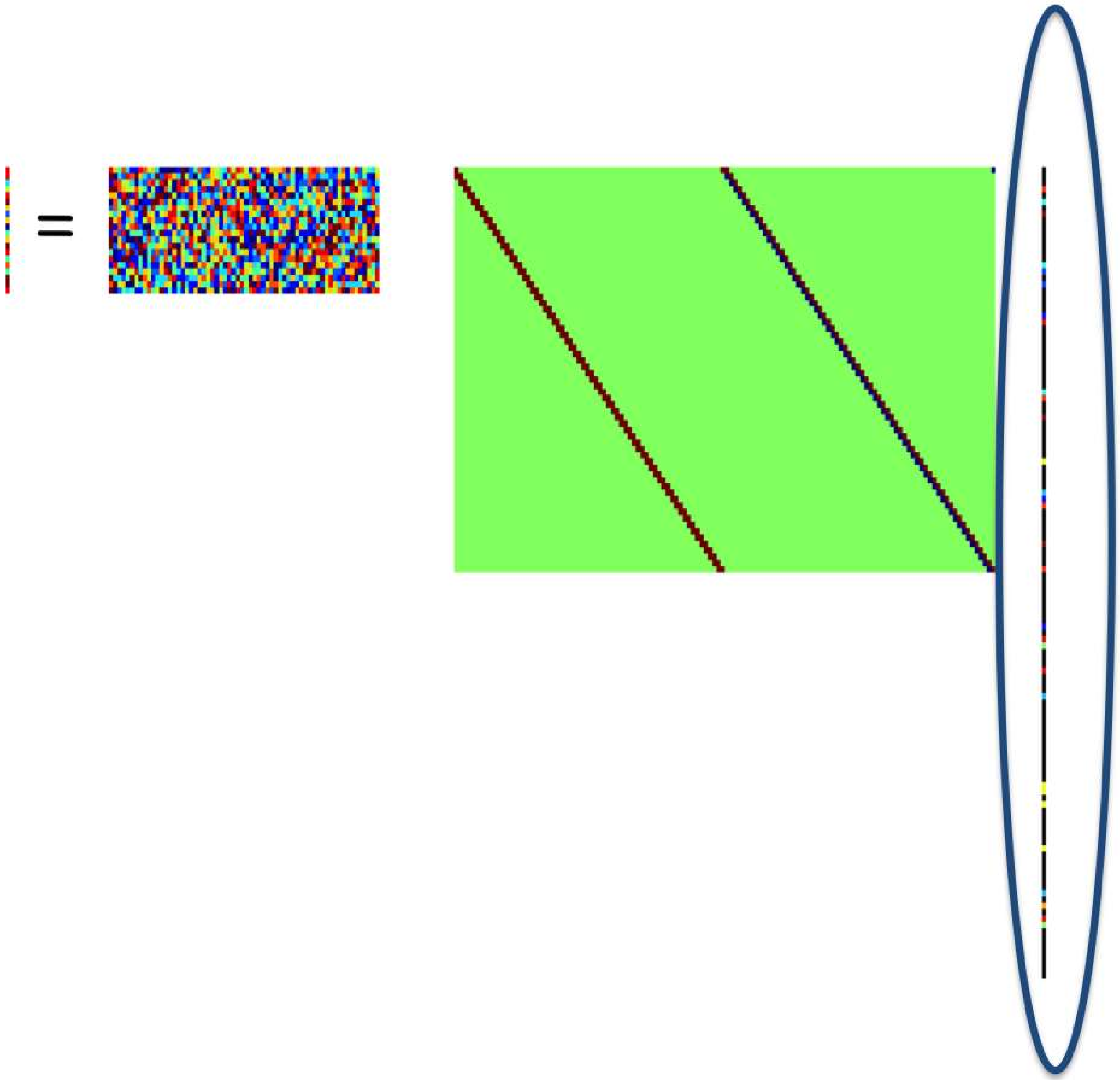
Visually



Visually



Visually



Synthesis basis pursuit for coefficient/signal recovery

Synthesis basis pursuit for coefficient recovery

$$\hat{Z} := \operatorname*{argmin}_{\boldsymbol{z} \in \mathbb{R}^d} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_2 \leq \eta. \qquad (\text{BP}_\eta^{\text{coef}})$$

$\boldsymbol{D} \in \mathbb{R}^{n \times d}$

- when $n = d$, for instance $\boldsymbol{D} = \mathbf{Id}$ (or any B.O.S) $\rightsquigarrow$ classical basis pursuit can recover any s-sparse vector $\boldsymbol{z}_0$ w.h.p. if $\boldsymbol{A}$ is sub-Gaussian with

$$m \gtrsim s \cdot \log(2n/s)$$

Synthesis basis pursuit for coefficient/signal recovery

Synthesis basis pursuit for coefficient recovery

$$\hat{Z} := \operatorname*{argmin}_{\boldsymbol{z} \in \mathbb{R}^d} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_2 \leq \eta. \qquad (\mathrm{BP}_\eta^{\mathrm{coef}})$$

$\boldsymbol{D} \in \mathbb{R}^{n \times d}$

- when $n = d$, for instance $\boldsymbol{D} = \mathbf{Id}$ (or any B.O.S) $\rightsquigarrow$ classical basis pursuit can recover any s-sparse vector $\boldsymbol{z}_0$ w.h.p. if $\boldsymbol{A}$ is sub-Gaussian with

$$m \gtrsim s \cdot \log(2n/s)$$

- in practice $n \ll d$, redundant $\boldsymbol{D}$

Synthesis basis pursuit for coefficient/signal recovery

Synthesis basis pursuit for coefficient recovery

$$\hat{Z} := \operatorname*{argmin}_{\boldsymbol{z}\in\mathbb{R}^d} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_2 \leq \eta. \qquad (\mathrm{BP}_\eta^{\mathrm{coef}})$$

$\boldsymbol{D} \in \mathbb{R}^{n\times d}$

- when $n = d$, for instance $\boldsymbol{D} = \mathbf{Id}$ (or any B.O.S) $\rightsquigarrow$ classical basis pursuit can recover any s-sparse vector $\boldsymbol{z}_0$ w.h.p. if

$$m \gtrsim s \cdot \log(2n/s).$$

- in practice $n \ll d$, redundant $\boldsymbol{D}$
  $\rightsquigarrow$ representations not necessarily unique
  $\rightsquigarrow$ can't expect to recover a specific representation via $(\mathrm{BP}_\eta^{\mathrm{coef}})$

One should be interested instead in:

Synthesis basis pursuit for signal recovery

$$\hat{X} := \boldsymbol{D} \cdot \underbrace{\left(\operatorname*{argmin}_{\boldsymbol{z}\in\mathbb{R}^d} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_2 \leq \eta\right)}_{=:\hat{Z}}. \qquad (\mathrm{BP}_\eta^{\mathrm{sig}})$$

Synthesis basis pursuit for coefficient/signal recovery

In the noiseless case (i.e., when $\boldsymbol{e} = \mathbf{0}$ and $\eta = 0$),

- it might be the case that $\hat{Z} \neq \{\boldsymbol{z}_0\}$ (coefficient recovery fails)
- but hope that $\hat{X} = \boldsymbol{D} \cdot \hat{Z} = \{\boldsymbol{x}_0\}$ (signal recovery successes)

Synthesis basis pursuit for coefficient/signal recovery

In the noiseless case (i.e., when $\boldsymbol{e} = \mathbf{0}$ and $\eta = 0$),

- it might be the case that $\hat{Z} \neq \{\boldsymbol{z}_0\}$ (coefficient recovery fails)
- but hope that $\hat{X} = \boldsymbol{D} \cdot \hat{Z} = \{\boldsymbol{x}_0\}$ (signal recovery successes)

Questions

(Q1) When coefficient recovery ≠ signal recovery?

(Q2) How many measurements are required for coefficient recovery? signal recovery?

(Q3) In case of coefficient and signal recovery, what about robustness to measurement noise?

Related works on the synthesis formulation

[Rauhut, Schnass and Vandergheynst 2008]

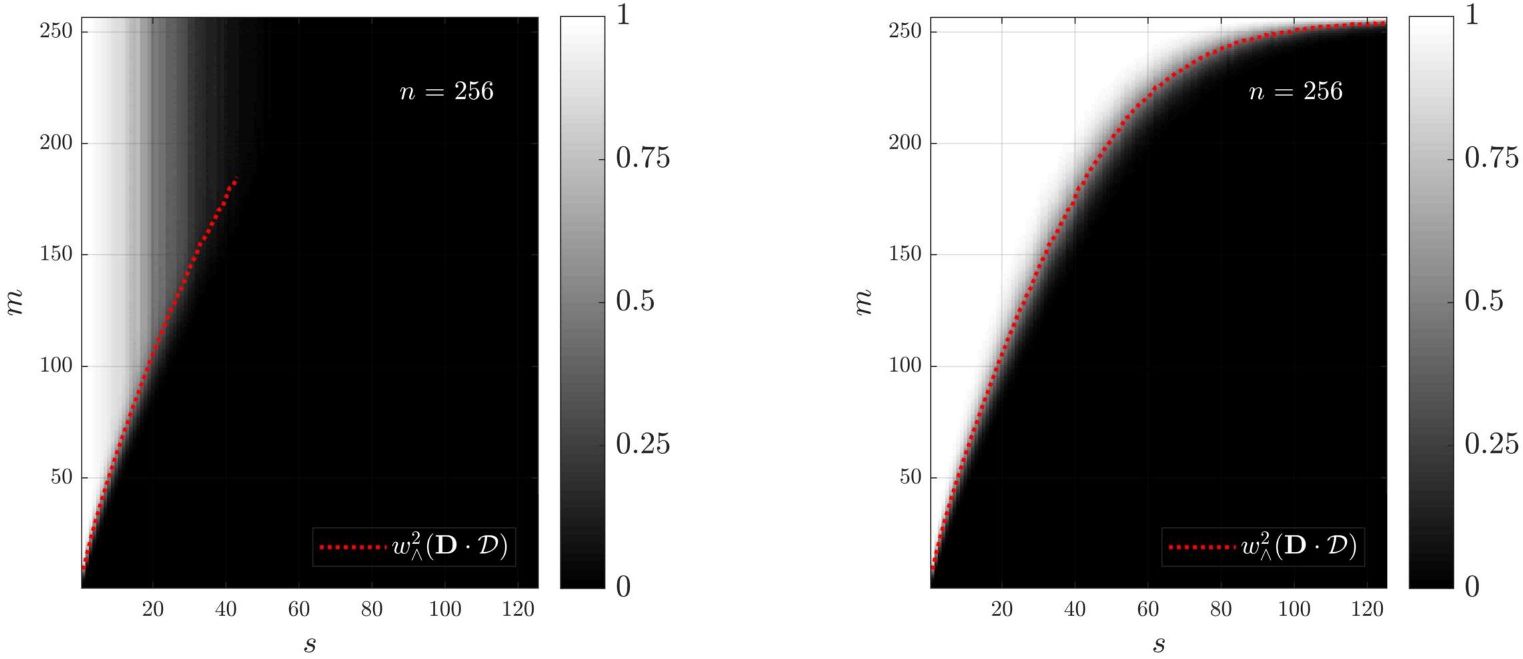
[Casazza, Chen, and Lynch, 2019]

✗ Address the coefficient recovery and not the signal one

[Rauhut, Schnass and Vandergheynst 2008]

[Casazza, Chen, and Lynch, 2019]

✗ Address the coefficient recovery and not the signal one



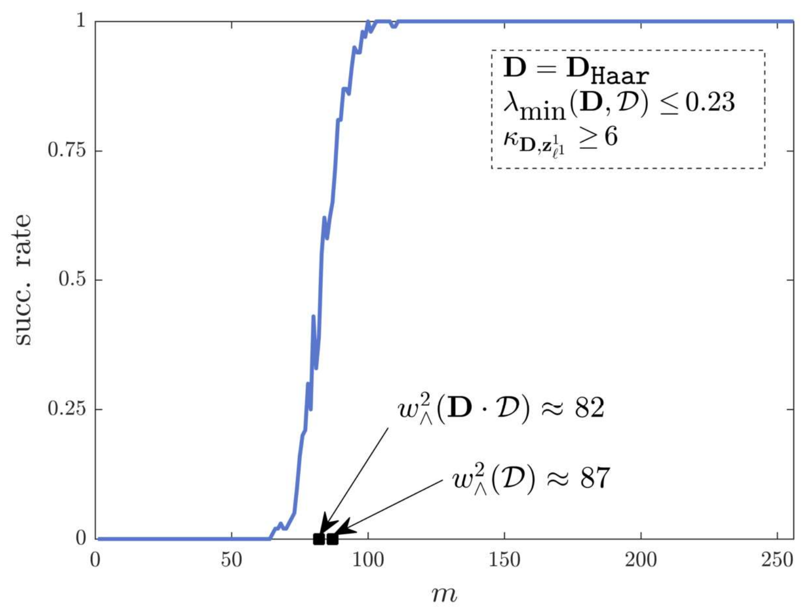
Phase transitions of coefficient and signal recovery by ℓ^1-synthesis.

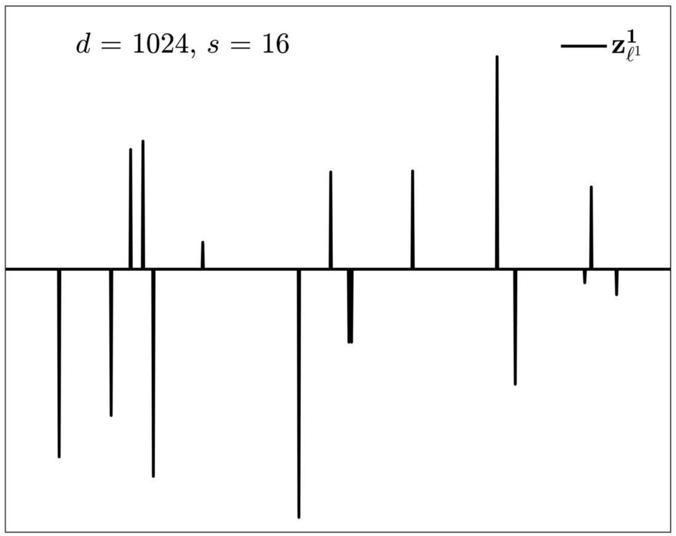
Related works on the synthesis formulation

[Rauhut, Schnass and Vandergheynst 2008]

[Casazza, Chen, and Lynch, 2019]

- ✗ Address the coefficient recovery and not the signal one
- ✗ Uniform results over all *s*-sparse coefficient vectors





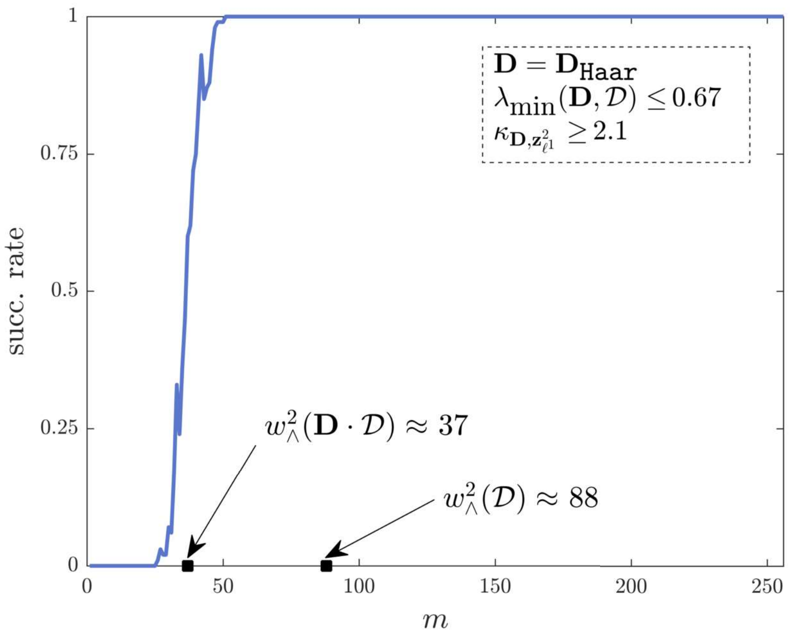
Related works on the synthesis formulation

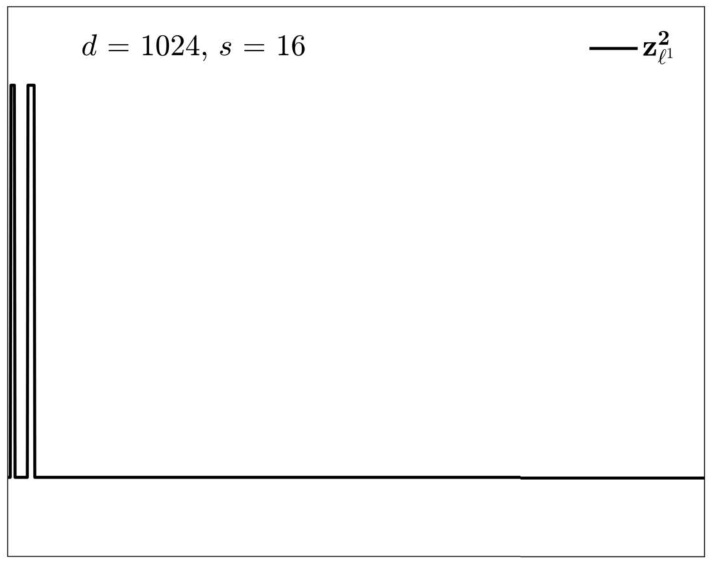
[Rauhut, Schnass and Vandergheynst 2008]

[Casazza, Chen, and Lynch, 2019]

✗ Address the coefficient recovery and not the signal one

✗ Uniform results over all *s*-sparse coefficient vectors





Related works on the synthesis formulation

[Rauhut, Schnass and Vandergheynst 2008]

[Casazza, Chen, and Lynch, 2019]

- ✗ Address the coefficient recovery and not the signal one
- ✗ Uniform results over all s-sparse coefficient vectors
- ✗ Rely on strong assumptions on $\boldsymbol{D}$: RIP, NSP, incoherence ...
- ✗ Forget about redundant representation systems $\rightsquigarrow$ highly coherent and with many linear dependencies
- ✗ Square-root bottleneck: The Welch bound reveals that incoherence can only be satisfied for $s \lesssim \sqrt{n}$.

Related works on the synthesis formulation

[Rauhut, Schnass and Vandergheynst 2008]

[Casazza, Chen, and Lynch, 2019]

- ✗ Address the coefficient recovery and not the signal one
- ✗ Uniform results over all s-sparse coefficient vectors
- ✗ Rely on strong assumptions on $\boldsymbol{D}$: RIP, NSP, incoherence ...
- ✗ Forget about redundant representation systems $\rightsquigarrow$ highly coherent and with many linear dependencies
- ✗ Square-root bottleneck: The Welch bound reveals that incoherence can only be satisfied for $s \lesssim \sqrt{n}$.

Our goals

- Need for local and non-uniform approach: signal-dependent analysis is crucial for redundant representation systems
- Avoiding strong assumptions on the dictionary
- Distinguishing signal and coefficient recovery

Summary

A key quantity: Minimum conic singular value

Consider the generalized basis pursuit

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \eta, \tag{BP_η^f}$$

$f : \mathbb{R}^n \to \mathbb{R}$ is convex, supposed to reflect the "low complexity" of the signal $\boldsymbol{x}_0$.

[Chandrasekaran et al. 2012, Tropp 2015]

A deterministic error bound for (BP_η^f)

(a) If $\eta = 0$,

exact recovery of $\boldsymbol{x}_0$ by solving $\mathrm{BP}_{\eta=0}^f \iff \lambda_{\min}(\boldsymbol{A}; \mathcal{D}_\wedge(f, \boldsymbol{x}_0)) > 0$

(b) In addition, any solution $\hat{\boldsymbol{x}}$ of (BP_η^f) satisfies

$$\left\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\right\|_2 \leq \frac{2\eta}{\lambda_{\min}(\boldsymbol{A}; \mathcal{D}_\wedge(f, \boldsymbol{x}_0))}. \tag{2}$$

A key quantity: Minimum conic singular value

Consider the generalized basis pursuit

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \le \eta, \tag{BP_η^f}$$

$f : \mathbb{R}^n \to \mathbb{R}$ is convex, supposed to reflect the "low complexity" of the signal $\boldsymbol{x}_0$.

[Chandrasekaran et al. 2012, Tropp 2015]

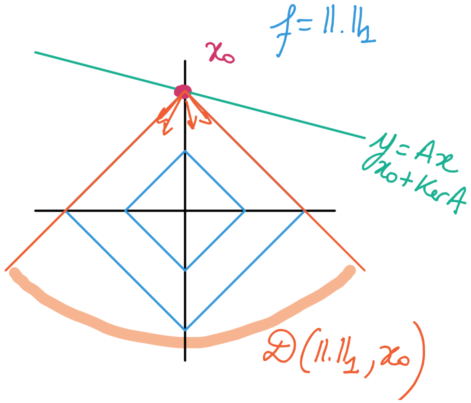
A deterministic error bound for (BP_η^f)

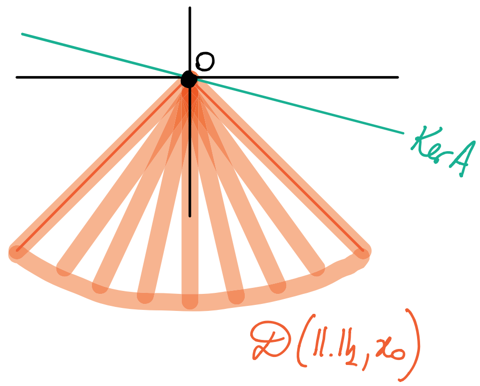
(a) If $\eta = 0$,

exact recovery of $\boldsymbol{x}_0$ by solving $\mathrm{BP}_{\eta=0}^f$ $\iff$ $\lambda_{\min}\left(\boldsymbol{A}; \mathcal{D}_\wedge(f, \boldsymbol{x}_0)\right) > 0$

(b) In addition, any solution $\hat{\boldsymbol{x}}$ of (BP_η^f) satisfies

$$\left\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\right\|_2 \le \frac{2\eta}{\lambda_{\min}\left(\boldsymbol{A}; \mathcal{D}_\wedge(f, \boldsymbol{x}_0)\right)}. \tag{2}$$





A key quantity: Minimum conic singular value

Consider the generalized basis pursuit

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \eta, \qquad (\mathrm{BP}^f_\eta)$$

$f : \mathbb{R}^n \to \mathbb{R}$ is convex, supposed to reflect the "low complexity" of the signal $\boldsymbol{x}_0$.

[Chandrasekaran et al. 2012, Tropp 2015]

A deterministic error bound for (BP^f_η)

(a) If $\eta = 0$,

exact recovery of $\boldsymbol{x}_0$ by solving $\mathrm{BP}^f_{\eta=0}$ $\iff$ $\lambda_{\min}\left(\boldsymbol{A}; \mathcal{D}_\wedge(f, \boldsymbol{x}_0)\right) > 0$

(b) In addition, any solution $\hat{\boldsymbol{x}}$ of (BP^f_η) satisfies

$$\left\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\right\|_2 \leq \frac{2\eta}{\lambda_{\min}\left(\boldsymbol{A}; \mathcal{D}_\wedge(f, \boldsymbol{x}_0)\right)}. \qquad (2)$$

- $\lambda_{\min}\left(\boldsymbol{A}; \mathcal{D}_\wedge(f, \boldsymbol{x}_0)\right)$ can be NP-hard to compute
- But there exists an estimate in the sub-Gaussian case!
- Through the Gordon's Escape Through a Mesh theorem

Control of $\lambda_{\min}(\boldsymbol{A}; \mathcal{D}_\wedge(f, \boldsymbol{x}_0))$ by the conic mean width $w_\wedge(\mathcal{D}(f, \boldsymbol{x}_0))$

Let $K \subseteq \mathbb{R}^n$ be a set. For $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Id})$ a standard Gaussian random vector,

(a) The **(global) mean width** of K is defined as $w(K) := \mathbb{E}\left[\sup_{\boldsymbol{h} \in K} \langle \boldsymbol{g}, \boldsymbol{h} \rangle\right]$.

(b) The **conic mean width** of K is given by $w_\wedge(K) := w(\mathrm{cone}(K) \cap \mathcal{S}^{n-1})$.

[Amelunxen, Lotz, McCoy, Tropp (2014)]

Sharp phase transition

In the noiseless case, $\mathrm{BP}^f_{\eta=0}$:

fails w.h.p. when

$$m \lesssim w_\wedge^2(\mathcal{D}(f, \boldsymbol{x}_0))$$

succeeds w.h.p. when

$$m \gtrsim w_\wedge^2(\mathcal{D}(f, \boldsymbol{x}_0))$$

Take-home messages on the generalized BP

- Robust signal recovery via the generalized basis pursuit (BP^f_η) is characterized by the minimum conic singular value $\lambda_{\min}(\boldsymbol{A}; \mathcal{D}_\wedge(f, \boldsymbol{x}_0))$.
- The required number of sub-Gaussian random measurements can be determined by the conic mean width of f at $\boldsymbol{x}_0$ $w_\wedge^2(\mathcal{D}(f, \boldsymbol{x}_0))$.
- $w_\wedge^2(\mathcal{D}(f, \boldsymbol{x}_0))$ gives a phase transition for the recovery success via (BP^f_η), in the noiseless case.

Summary

Sampling rate for coefficient recovery

$$\hat{Z} := \operatorname*{argmin}_{\boldsymbol{z}\in\mathbb{R}^d} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_2 \leq \eta. \qquad (\mathrm{BP}_\eta^{\mathrm{coef}})$$

Need to control $\lambda_{\min}\left(\boldsymbol{A}\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z})\right)$

Theorem (Coefficient recovery)

Let $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ be a dictionary and $\boldsymbol{z}_{\ell^1} \in \mathbb{R}^d$ such that $\boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{z}_{\ell^1} \in \mathbb{R}^n$, be the unique representer of $\boldsymbol{x}_0$ of minimal ℓ^1-norm

$$\lambda_{min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})\right) > 0.$$

Sampling rate for coefficient recovery

$$\hat{Z} := \operatorname*{argmin}_{\boldsymbol{z}\in\mathbb{R}^d} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_2 \leq \eta. \tag{$\mathrm{BP}_\eta^{\mathrm{coef}}$}$$

Need to control $\lambda_{\min}\left(\boldsymbol{A}\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z})\right)$

Theorem (Coefficient recovery)

Let $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ *be a dictionary and* $\boldsymbol{z}_{\ell^1} \in \mathbb{R}^d$ *such that* $\boldsymbol{x}_0 = \boldsymbol{D}\boldsymbol{z}_{\ell^1} \in \mathbb{R}^n$, *be the unique representer of* $\boldsymbol{x}_0$ *of minimal* ℓ^1*-norm*

$$\lambda_{min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})\right) > 0.$$

Then $\forall u > 0$, *with probability* $\geq 1 - e^{-u^2/2}$: *if*

$$m > m_0 := \left(w_\wedge\left(\boldsymbol{D}\cdot\mathcal{D}(\|\cdot\|_1; \boldsymbol{z}_{\ell^1})\right) + u\right)^2 + 1, \tag{3}$$

then any solution $\hat{\boldsymbol{z}}$ *to the program* ($\mathrm{BP}_\eta^{\mathrm{coef}}$) *satisfies*

$$\left\|\boldsymbol{z}_{\ell^1} - \hat{\boldsymbol{z}}\right\|_2 \leq \frac{2\eta}{\lambda_{min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1; \boldsymbol{z}_{\ell^1})\right)\cdot\left(\sqrt{m-1} - \sqrt{m_0 - 1}\right)}. \tag{4}$$

$$\lambda_{\min}\left(\boldsymbol{A}\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})\right) > \lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})\right)\cdot\inf\left\{\|\boldsymbol{A}\boldsymbol{x}\|_2 : \boldsymbol{x}\in\boldsymbol{D}\mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})\cap\mathcal{S}^{n-1}\right\}$$

Sampling rate for coefficient recovery

(a) No assumption on the dictionary $\boldsymbol{D}$ and the coefficient representation $\boldsymbol{z}_{\ell^1}$, except for

$$\lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}_{\wedge}(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})\right) > 0$$

which is

- a necessary condition for the theorem to hold true
- involved to ensure

(b) $w_{\wedge}^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1; \boldsymbol{z}_{\ell^1}))$ drives the sampling rate for coefficient recovery by ($\mathrm{BP}_{\eta}^{\mathrm{coef}}$).

(c) Lastly, the error bound shows that coefficient recovery is robust to measurement noise, provided that $\lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}_{\wedge}(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})\right) \gg 0$;

Convex gauge for signal recovery

Recall: synthesis basis pursuit for signal recovery

$$\hat{X} := \boldsymbol{D} \cdot \left(\operatorname*{argmin}_{\boldsymbol{z} \in \mathbb{R}^d} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_2 \leq \eta \right). \qquad (\mathrm{BP}_{\eta}^{\mathrm{sig}})$$

Convex gauge for signal recovery

Recall: synthesis basis pursuit for signal recovery

$$\hat{X} := \boldsymbol{D} \cdot \left(\operatorname*{argmin}_{\boldsymbol{z} \in \mathbb{R}^d} \|\boldsymbol{z}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}\|_2 \leq \eta \right). \qquad (\mathrm{BP}_\eta^{\mathrm{sig}})$$

Lemma (Gauge formulation)

Assume that $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{e}$, *with* $\|\boldsymbol{e}\|_2 \leq \eta$. *Let* $\boldsymbol{D} \in \mathbb{R}^{n \times d}$ *be a dictionary. Then,*

$$\hat{X} = \operatorname*{argmin}_{\boldsymbol{x} \in \mathbb{R}^n} p_{\boldsymbol{D} \cdot B_1^d}(\boldsymbol{x}) \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \eta.$$

Lemma (Descent cone)

Let $\boldsymbol{x}_0 \in \operatorname{ran}(\boldsymbol{D})$. *For* <u>*any*</u> $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$ *(*ℓ^1*-representers of* $\boldsymbol{x}_0$ *in* $\boldsymbol{D}$*),*

$$\mathcal{D}_\wedge(p_{\boldsymbol{D} \cdot B_1^d}, \boldsymbol{x}_0) = \boldsymbol{D} \cdot \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1}) \quad \text{and} \quad \mathcal{D}(p_{\boldsymbol{D} \cdot B_1^d}, \boldsymbol{x}_0) = \boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1, \boldsymbol{z}_{\ell^1}).$$

Sampling rate for signal recovery

Theorem (Signal recovery)

Let $\boldsymbol{D} \in \mathbb{R}^{n \times d}$ be a dictionary with $\boldsymbol{x}_0 \in \operatorname{ran}(\boldsymbol{D})$ and pick any $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$.
$\forall u > 0$, with probability $\geq 1 - e^{-u^2/2}$: if

$$m > m_0 := \left(w_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1 ; \boldsymbol{z}_{\ell^1})) + u \right)^2 + 1, \tag{5}$$

then any solution $\hat{\boldsymbol{x}}$ to the program ($\mathrm{BP}^{\mathrm{sig}}_{\eta=0}$) satisfies

$$\left\| \boldsymbol{x}_0 - \hat{\boldsymbol{x}} \right\|_2 \leq \frac{2\eta}{\sqrt{m-1} - \sqrt{m_0 - 1}}. \tag{6}$$

Sampling rate for signal recovery

Theorem (Signal recovery)

Let $\boldsymbol{D} \in \mathbb{R}^{n \times d}$ *be a dictionary with* $\boldsymbol{x}_0 \in \operatorname{ran}(\boldsymbol{D})$ *and pick* <u>*any*</u> $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$.
$\forall u > 0$, *with probability* $\geq 1 - e^{-u^2/2}$: *if*

$$m > m_0 := \left(w_\wedge\left(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1 ; \boldsymbol{z}_{\ell^1})\right) + u\right)^2 + 1, \tag{5}$$

then any solution $\hat{\boldsymbol{x}}$ *to the program* ($\mathrm{BP}^{\mathrm{sig}}_{\eta=0}$) *satisfies*

$$\left\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\right\|_2 \leq \frac{2\eta}{\sqrt{m-1} - \sqrt{m_0 - 1}}. \tag{6}$$

(a) Again, $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1 ; \boldsymbol{z}_{\ell^1}))$ drives the sampling rate

Sampling rate for signal recovery

Theorem (Signal recovery)

Let $\boldsymbol{D} \in \mathbb{R}^{n \times d}$ *be a dictionary with* $\boldsymbol{x}_0 \in \operatorname{ran}(\boldsymbol{D})$ *and pick* <u>*any*</u> $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$.
$\forall u > 0$, *with probability* $\geq 1 - e^{-u^2/2}$: *if*

$$m > m_0 := \left(w_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1 ; \boldsymbol{z}_{\ell^1})) + u\right)^2 + 1, \tag{5}$$

then any solution $\hat{\boldsymbol{x}}$ *to the program* ($\mathrm{BP}^{\mathrm{sig}}_{\eta=0}$) *satisfies*

$$\left\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\right\|_2 \leq \frac{2\eta}{\sqrt{m-1} - \sqrt{m_0 - 1}}. \tag{6}$$

(a) Again, $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1 ; \boldsymbol{z}_{\ell^1}))$ drives the sampling rate

(b) But the set of minimal ℓ^1-representers is not required to be a singleton: The descent cone in the signal space may be evaluated at any possible $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$.

Sampling rate for signal recovery

Theorem (Signal recovery)

Let $\boldsymbol{D} \in \mathbb{R}^{n \times d}$ *be a dictionary with* $\boldsymbol{x}_0 \in \mathrm{ran}(\boldsymbol{D})$ *and pick* <u>*any*</u> $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$.
$\forall u > 0$*, with probability* $\geq 1 - e^{-u^2/2}$ *: if*

$$m > m_0 := \left(w_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}_{\ell^1})) + u\right)^2 + 1, \tag{5}$$

then any solution $\hat{\boldsymbol{x}}$ *to the program* ($\mathrm{BP}^{\mathrm{sig}}_{\eta=0}$) *satisfies*

$$\left\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\right\|_2 \leq \frac{2\eta}{\sqrt{m-1} - \sqrt{m_0 - 1}}. \tag{6}$$

(a) Again, $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}_{\ell^1}))$ drives the sampling rate

(b) But the set of minimal ℓ^1-representers is not required to be a singleton: The descent cone in the signal space may be evaluated at any possible $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$.

(c) Phase transition of signal recovery at m_0.

Sampling rate for signal recovery

Theorem (Signal recovery)

Let $\boldsymbol{D} \in \mathbb{R}^{n \times d}$ *be a dictionary with* $\boldsymbol{x}_0 \in \operatorname{ran}(\boldsymbol{D})$ *and pick any* $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$.
$\forall u > 0$*, with probability* $\geq 1 - e^{-u^2/2}$ *: if*

$$m > m_0 := (w_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1 ; \boldsymbol{z}_{\ell^1})) + u)^2 + 1, \tag{5}$$

then any solution $\hat{\boldsymbol{x}}$ *to the program* ($\mathrm{BP}^{\mathrm{sig}}_{\eta=0}$) *satisfies*

$$\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\|_2 \leq \frac{2\eta}{\sqrt{m-1} - \sqrt{m_0 - 1}}. \tag{6}$$

(a) Again, $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1 ; \boldsymbol{z}_{\ell^1}))$ drives the sampling rate

(b) But the set of minimal ℓ^1-representers is not required to be a singleton: The descent cone in the signal space may be evaluated at any possible $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$.

(c) Phase transition of signal recovery at m_0.

(d) No minimal conic singular value involved! (even 0 is allowed!)
$\rightsquigarrow$ In the case of simultaneous coefficient and signal recovery, the robustness to noise of ($\mathrm{BP}^{\mathrm{coef}}_{\eta}$) and ($\mathrm{BP}^{\mathrm{sig}}_{\eta=0}$) might still be different.

Summary

How to evaluate the conic mean width $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}_{\ell^1}))$?

- ✓ Tight and informative upper bounds for simple dictionaries such as orthogonal matrices
- ✗ Involved for general, possibly redundant transforms
- ✗ We cannot use classical argument based on polarity Indeed,
- ✗ A bound based on a local condition number is too pessimistic

$$w_\wedge^2(\underbrace{\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}_{\ell^1})}_{=:C}) \leq \frac{\|\boldsymbol{D}\|_2}{\lambda_{\min}\left(\boldsymbol{D}; \mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}_{\ell^1})\right)} \cdot \left(w_\wedge^2(\mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}_{\ell^1})) + 1\right)$$

A geometric bound instead

1. Decompose the cone into its lineality and its range $C = C_L \oplus C_R$

$$w_\wedge^2(C) \quad \lesssim \quad w_\wedge^2(C_L) + w_\wedge^2(C_R) + 1$$

2. The lineality C_L is the largest subspace contained in the cone, so $w_\wedge^2(C_L) \simeq \dim(C_L)$
3. The range is finitely generated, line-free, and contained into a circular cone of circumangle $\alpha < \pi/2$

$\rightsquigarrow$ new bound on the conic mean width for such cones

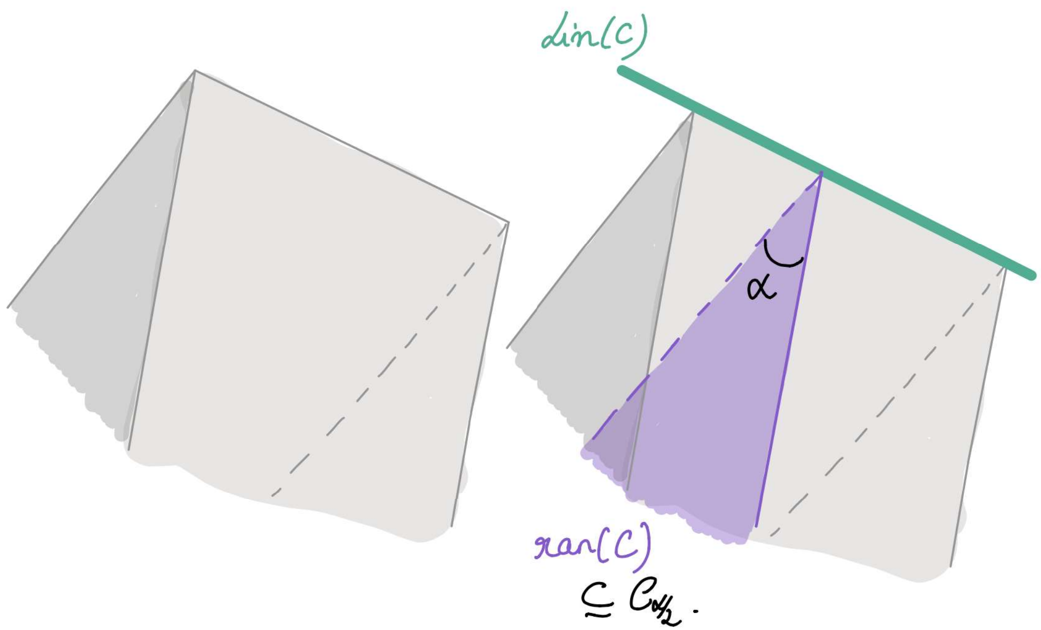
A geometric bound instead

1. Decompose the cone into its lineality and its range $C = C_L \oplus C_R$

$$w_\wedge^2(C) \quad \lesssim \quad w_\wedge^2(C_L) + w_\wedge^2(C_R) + 1$$

2. The lineality C_L is the largest subspace contained in the cone, so $w_\wedge^2(C_L) \simeq \dim(C_L)$
3. The range is finitely generated, line-free, and contained into a circular cone of circumangle $\alpha < \pi/2$

   $\rightsquigarrow$ new bound on the conic mean width for such cones



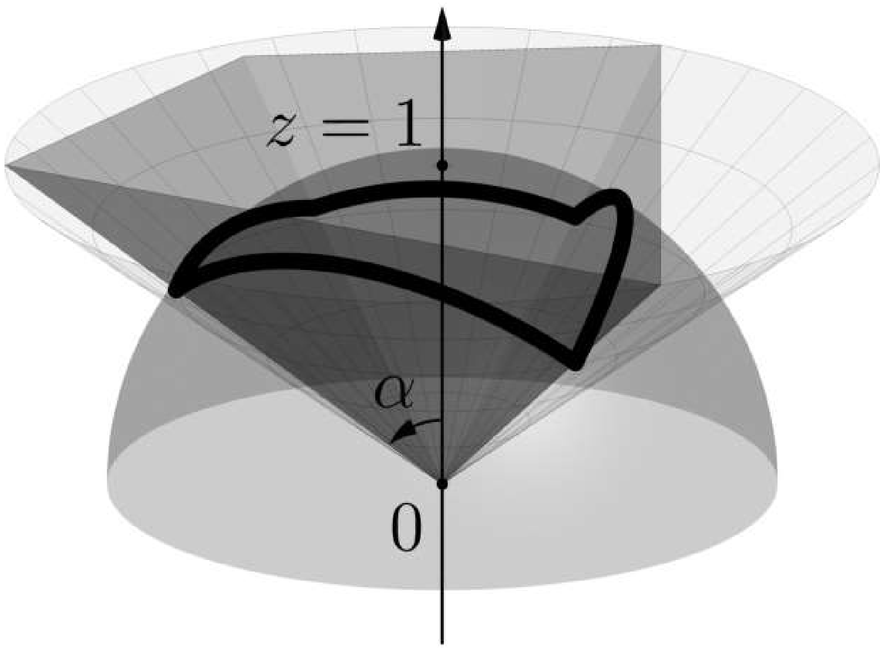
A geometric bound instead

1. Decompose the cone into its lineality and its range $C = C_L \oplus C_R$

$$w_\wedge^2(C) \quad \lesssim \quad w_\wedge^2(C_L) + w_\wedge^2(C_R) + 1$$

2. The lineality C_L is the largest subspace contained in the cone, so $w_\wedge^2(C_L) \simeq \dim(C_L)$
3. The range is finitely generated, line-free, and contained into a circular cone of circumangle $\alpha < \pi/2$

   $\rightsquigarrow$ new bound on the conic mean width for such cones



Proposition

Let $\mathbf{D} \in \mathbb{R}^{n\times d}$ be a dictionary and let $\mathbf{x}_0 \in \mathrm{ran}(\mathbf{D}) \setminus \{\mathbf{0}\}$.

Let $C := \mathcal{D}_\wedge(p_{\mathbf{D}\cdot \mathrm{B}_1^d}, \mathbf{x}_0) = \mathbf{D} \cdot \mathcal{D}(\|\cdot\|_1\,; \mathbf{z}_{\ell^1})$ denote the descent cone of the gauge at $\mathbf{x}_0$.

Let $\mathbf{z}_{\ell^1} \in \mathrm{ri}(Z_{\ell^1})$ be any minimal ℓ^1-representer of $\mathbf{x}_0$ in $\mathbf{D}$ with maximal support and set $\bar{\mathcal{S}} = \mathrm{supp}(\mathbf{z}_{\ell^1})$ as well as $\bar{s} = \#\bar{\mathcal{S}}$.
Assume $\bar{s} < d$.

Then we have:

(a) The lineality space of C has a dimension less than $\bar{s} - 1$ and is given by

$$C_L = \mathrm{span}(\bar{s} \cdot \mathrm{sign}(z_{\ell^1,i}) \cdot \mathbf{d}_i - \mathbf{D} \cdot \mathrm{sign}(\mathbf{z}_{\ell^1}) : i \in \bar{\mathcal{S}}). \tag{7}$$

(b) The range of C is a $2(d - \bar{s})$-polyhedral α-cone given by:

$$C_R = \mathrm{cone}(\mathbf{r}_j^{\pm\perp} : j \in \bar{\mathcal{S}}^c) \text{ with } \mathbf{r}_j^{\pm\perp} := \mathbf{P}_{C_L^\perp}(\pm\bar{s} \cdot \mathbf{d}_j - \mathbf{D} \cdot \mathrm{sign}(\mathbf{z}_{\ell^1}))\,. \tag{8}$$

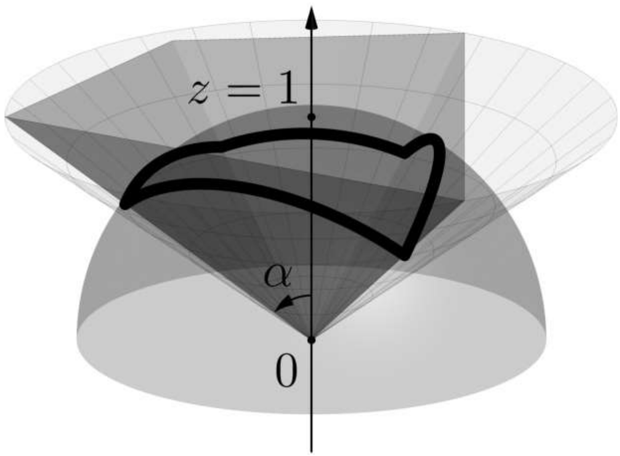
Circumangle for pointed polyhedral cone

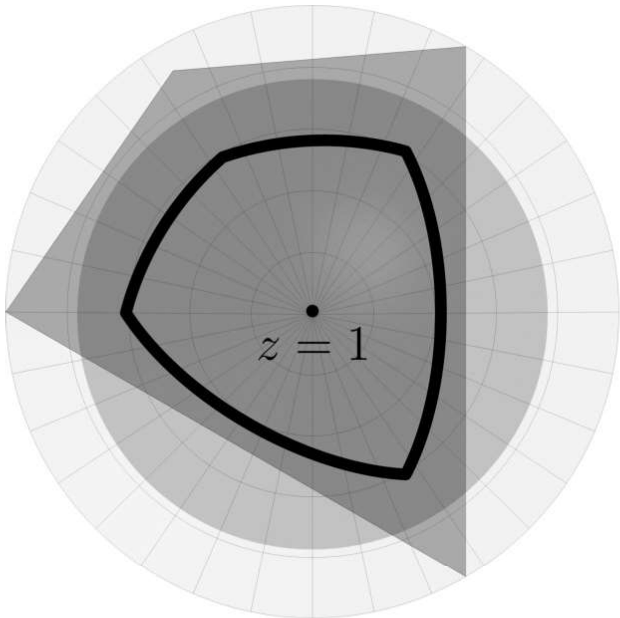
Proposition: Circumangle and circumcenter of polyhedral cones

Let $\boldsymbol{x}_i \in \mathcal{S}^{n-1}$ for $i \in [k]$ and let $C = \text{cone}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ be a nontrivial pointed polyhedral cone. Finding the circumcenter and circumangle of C amounts to solving the convex problem:

$$\cos(\alpha) = \sup_{\boldsymbol{\theta} \in \mathsf{B}_2^n} \inf_{i \in [k]} \langle \boldsymbol{\theta}, \boldsymbol{x}_i \rangle.$$

- ✓ possible to numerically compute the circumangle of pointed polyhedral cones.
- ≠ the minimum conic singular value is intractable in general





Cmw for k-polyhedral cone contained into α-circular cones

Proposition

For $k \geq 5$, the conic mean width of a k-polyhedral cone contained into an α-circular cone C in $\mathbb{R}^n$ is bounded by

$$W(\alpha, k, n) \leq \tan\alpha \cdot \left(\sqrt{2\log\left(k/\sqrt{2\pi}\right)} + \frac{1}{\sqrt{2\log\left(k/\sqrt{2\pi}\right)}} \right) + \frac{1}{\sqrt{2\pi}}.$$

- the bound does not depend on the ambient dimension n,
- $\neq$ in contrast to the conic width of a circular cone.

Consequence for the sampling rate

Theorem

If $\bar{s} \leq d - 3$, we obtain that

$$w_\wedge^2(\mathcal{D}_\wedge(p_{\boldsymbol{D}\cdot B_1^d}, \boldsymbol{x}_0)) \leq \bar{s} + \left(\tan \alpha \cdot \left(\sqrt{2 \log\left(\frac{2(d-\bar{s})}{\sqrt{2\pi}}\right)} + 1 \right) + \frac{1}{\sqrt{2\pi}} \right)^2,$$

Consequence for the sampling rate

Theorem

If $\bar{s} \leq d - 3$, we obtain that

$$w_{\wedge}^2(\mathcal{D}_{\wedge}(p_{\boldsymbol{D}\cdot B_1^d}, \boldsymbol{x}_0)) \leq \bar{s} + \left(\tan\alpha \cdot \left(\sqrt{2\log\left(\frac{2(d-\bar{s})}{\sqrt{2\pi}}\right)} + 1\right) + \frac{1}{\sqrt{2\pi}}\right)^2,$$

Corollary

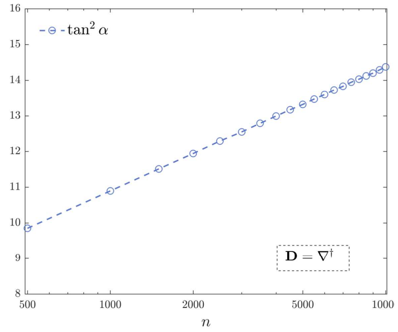
The critical number of measurements m_0 satisfies

$$m_0 \lesssim \bar{s} + \tan^2\alpha \cdot \log(2(d-\bar{s})/\sqrt{2\pi}). \tag{9}$$

The sampling rate is mainly governed by

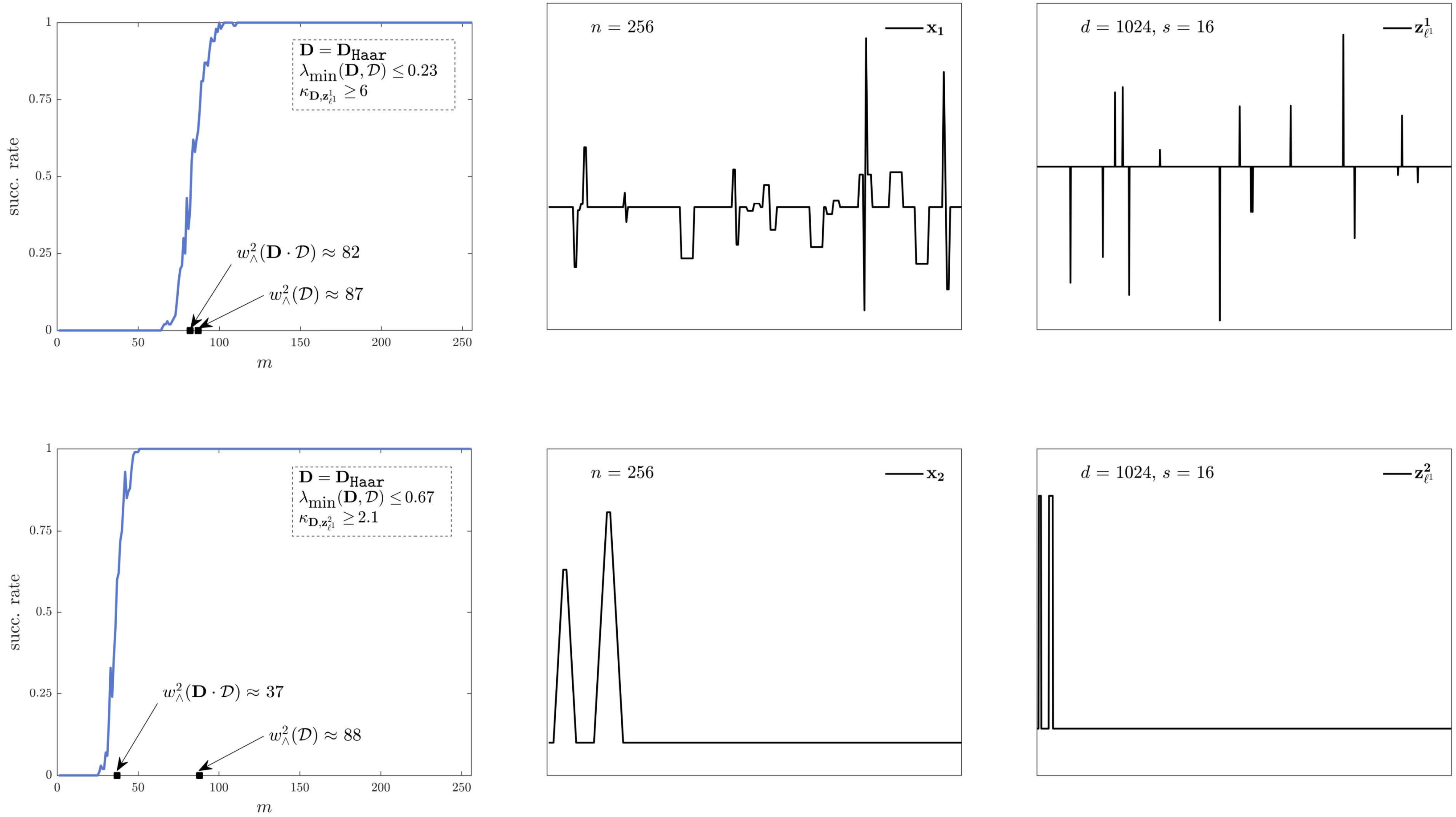
- the sparsity $\bar{s}$ of maximal support ℓ^1-representations of $\boldsymbol{x}_0$ in $\boldsymbol{D}$
- the “narrowness” of the remaining cone C_R, which is captured by its circumangle $\alpha \in [0, \pi/2)$
- The number of dictionary atoms only has a logarithmic influence.
  NB: comparable to the mean width of a convex polytope, which is mainly determined by its diameter and by the logarithm of its number of vertices.

Examples

$\boldsymbol{D}$	$\boldsymbol{x}_0 \in \mathbb{R}^n$	$m \gtrsim$	
$\boldsymbol{D} = \mathbf{Id} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	s-sparse vector	$2s\log(2(n-s)/\sqrt{2\pi})$	✓
Convolutional dictionary $\boldsymbol{D} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 \end{pmatrix}$	2-sparse $(1\,0\,0\ldots0\,1)^T$	$2 + 2\log(4n)$	(new) ✚
Total gradient variation $\boldsymbol{D} = \nabla^\dagger$	s-gradient sparse	Numerical evaluation $s \cdot \log^2(n)$ 	✓

Summary

Phase transition for coefficient recovery



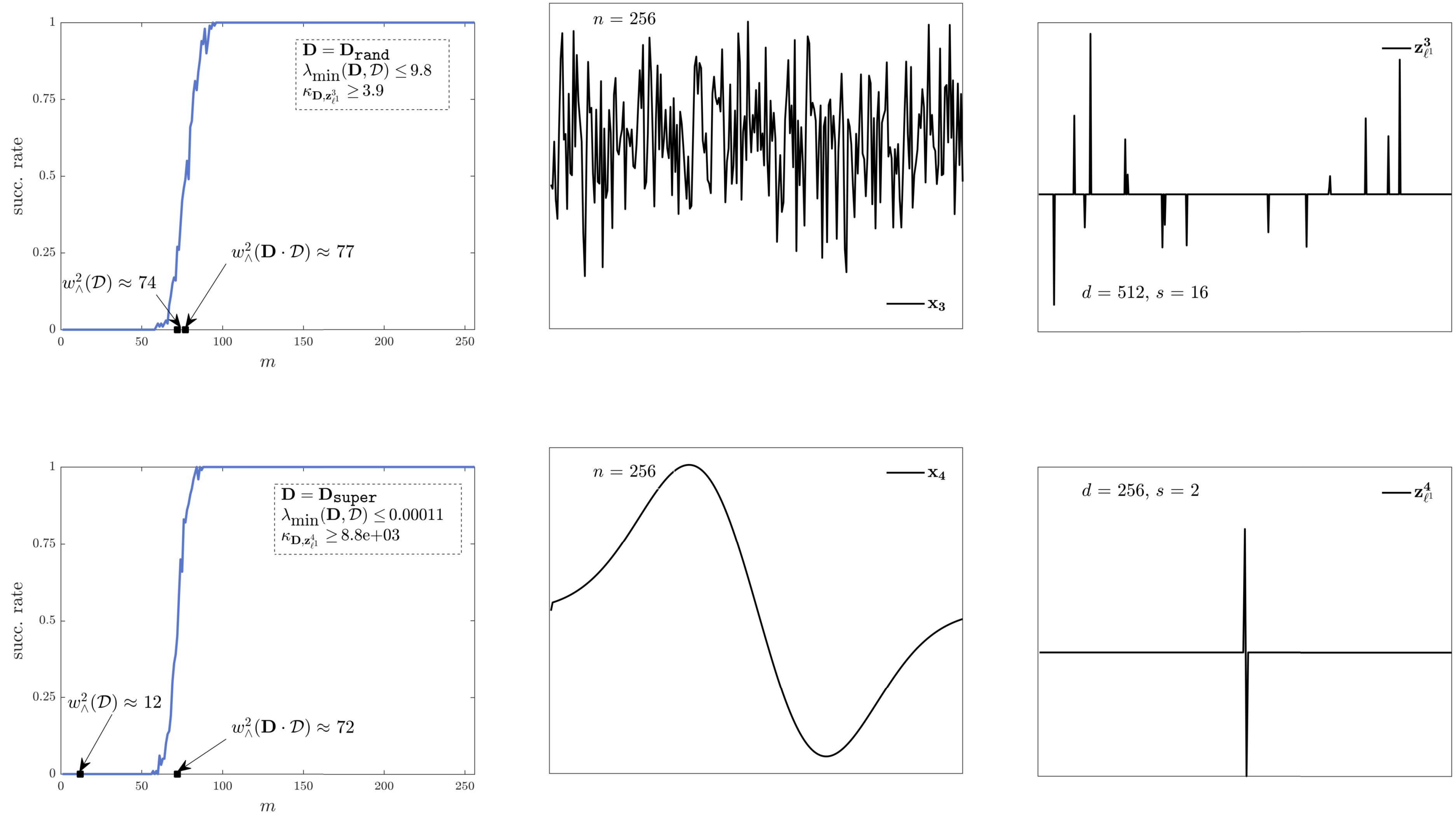
(i) ($\text{BP}^{\text{coef}}_{\eta=0}$) obeys a sharp phase transition in the number of measurements m

(ii) $w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}_{\ell^1}))$ accurately describes the sampling rate of ($\text{BP}^{\text{coef}}_{\eta=0}$)

(iii) Need of a non-uniform theory across the class of all s-sparse signals:

$$w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}^1_{\ell^1})) \neq w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}^2_{\ell^1})) \text{ and } w^2_\wedge(\mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}^1_{\ell^1})) = w^2_\wedge(\mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}^2_{\ell^1}))$$

Phase transition for coefficient recovery



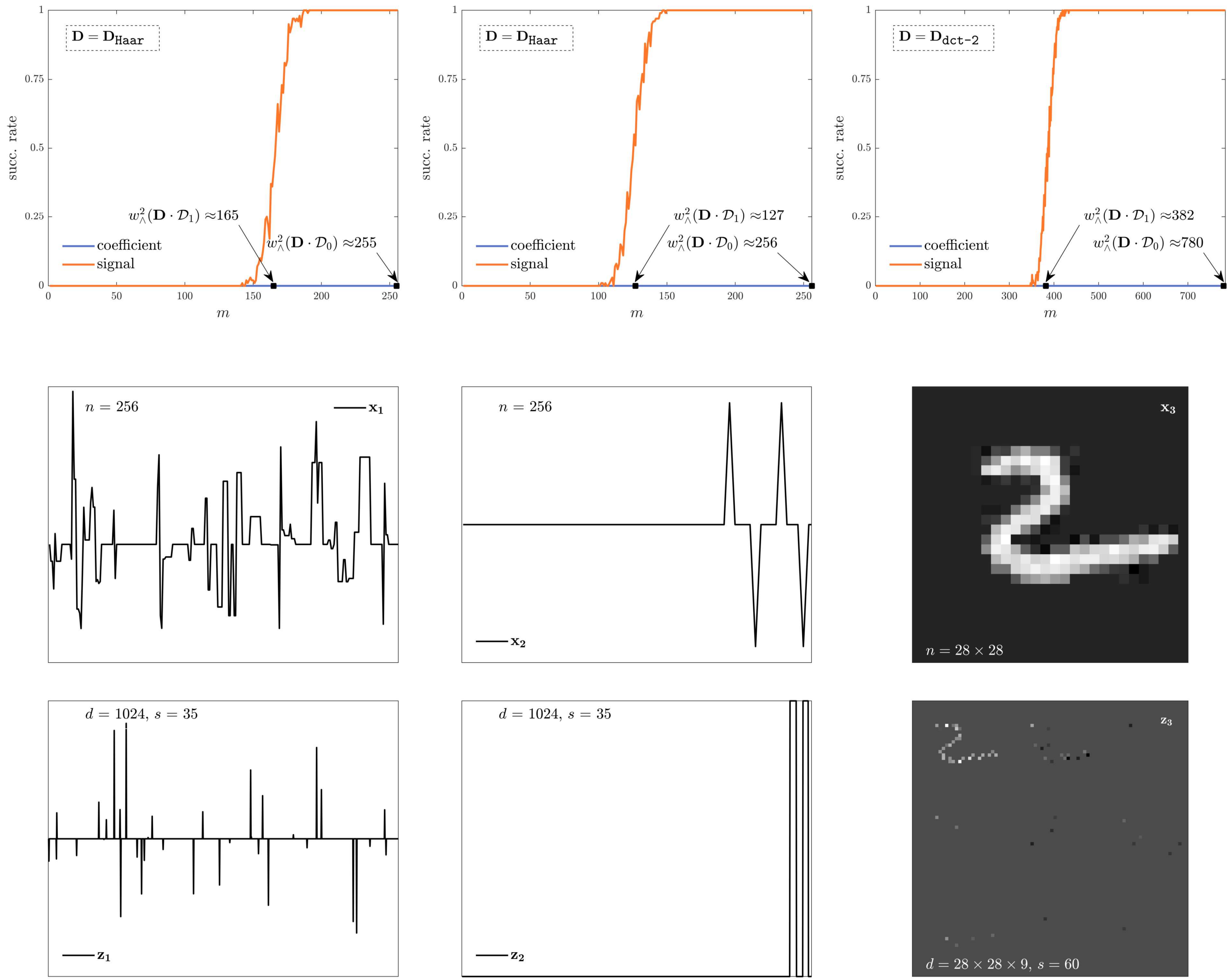
(iv) $w^2_\wedge(\mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}_{\ell^1}))$ does not describe the sampling rate of $(\mathrm{BP}^{\text{coef}}_{\eta=0})$. Indeed,

$$w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}^2_{\ell^1})) \ll \text{ or } \gg w^2_\wedge(\mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}^2_{\ell^1}))$$

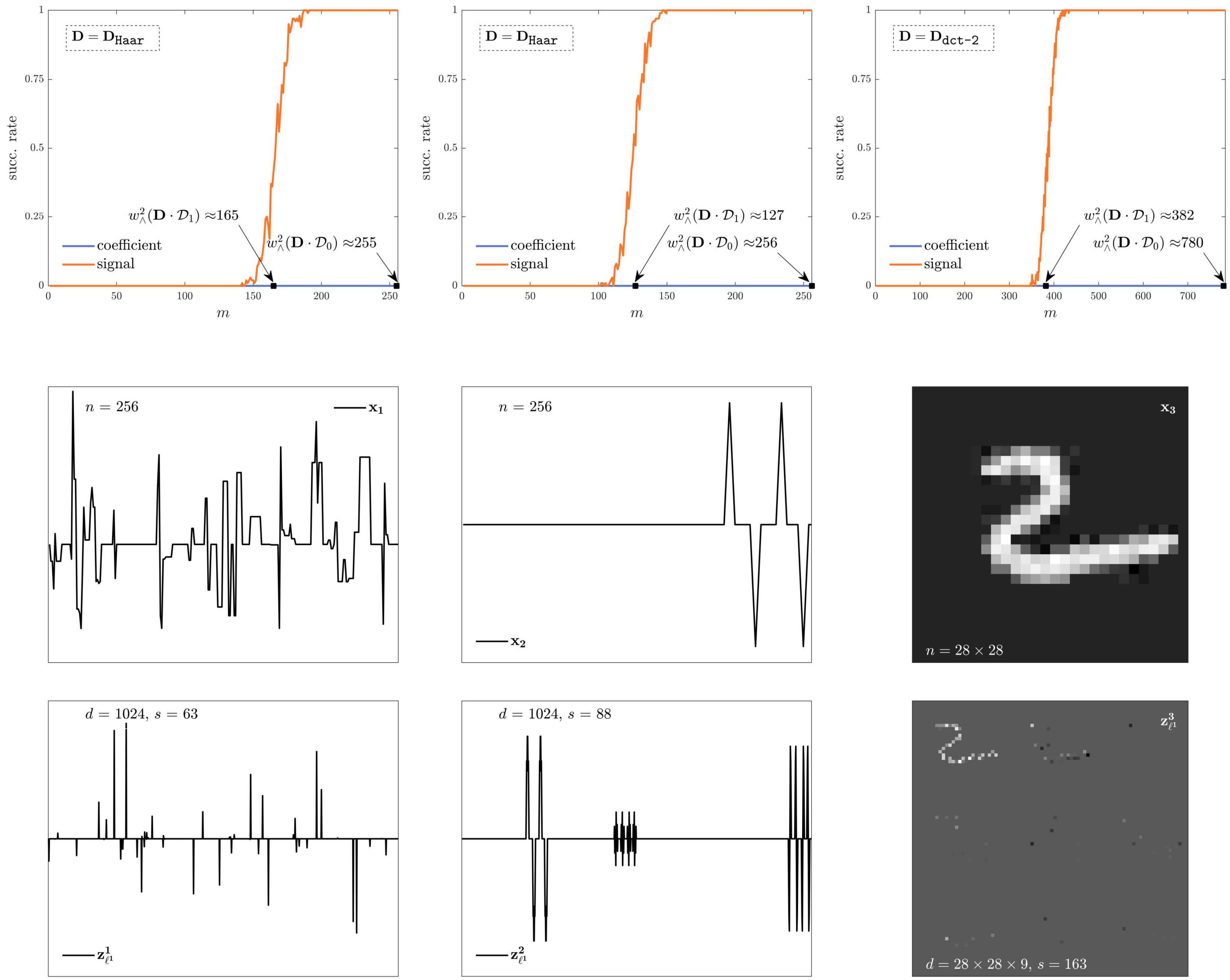
$\rightsquigarrow$ Sparsity alone is not a good proxy for the sampling complexity of $(\mathrm{BP}^{\text{coef}}_{\eta=0})$

(v) The local condition number $\kappa_{\boldsymbol{D},\boldsymbol{z}_{\ell^1}}$ might explode

Phase transition for signal recovery



Phase transition for signal recovery



Phase transition for signal recovery

(vi) ($\text{BP}^{\text{sig}}_{\eta=0}$) obeys a sharp phase transition in the number of measurements

However, a recovery of a coefficient representation via solving ($\text{BP}^{\text{coef}}_{\eta=0}$) is impossible in all three examples, even for $m = n$.

(vii) For any $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$, $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}_{\ell^1}))$ accurately describes the sampling rate of ($\text{BP}^{\text{sig}}_{\eta=0}$).

(vi) ($BP^{sig}_{\eta=0}$) obeys a sharp phase transition in the number of measurements

However, a recovery of a coefficient representation via solving ($BP^{coef}_{\eta=0}$) is impossible in all three examples, even for $m = n$.

(vii) For any $\boldsymbol{z}_{\ell^1} \in Z_{\ell^1}$, $w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}_{\ell^1}))$ accurately describes the sampling rate of ($BP^{sig}_{\eta=0}$).

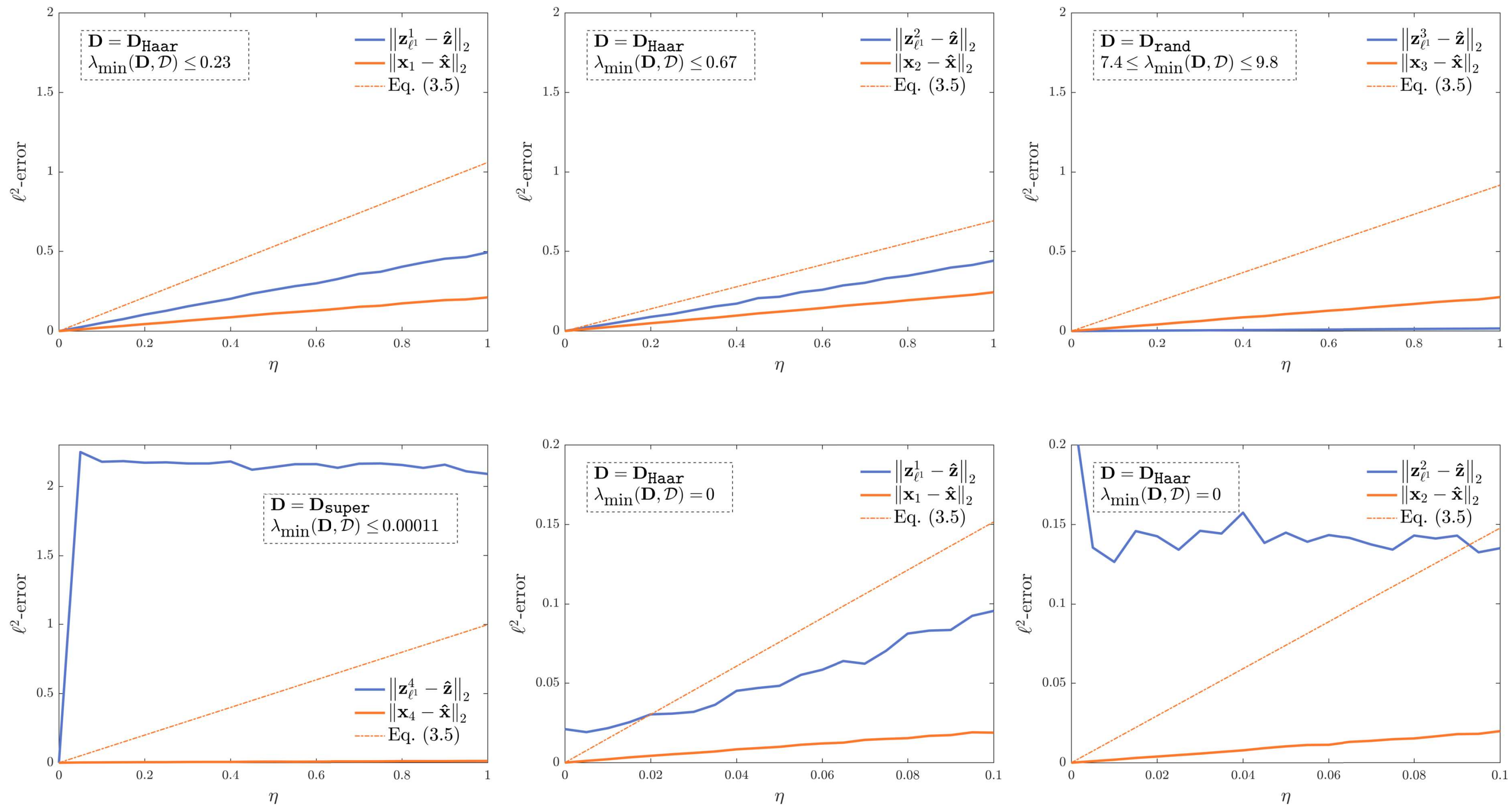
(viii) For any other sparse representation $\boldsymbol{z} \notin Z_{\ell^1}$, $w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}))$ does not describe the sampling rate of ($BP^{sig}_{\eta=0}$).

- Indeed, observe that we have $w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}_i)) \approx n$ in all three examples.
- $\|\boldsymbol{z}_1\|_0 = 35 = \|\boldsymbol{z}_2\|_0$, but different phase transitions locations
- Although $\|\boldsymbol{z}^1_{\ell^1}\|_0 < \|\boldsymbol{z}^2_{\ell^1}\|_0$, we have that

$$w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}^1_{\ell^1})) > w^2_\wedge(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,;\boldsymbol{z}^2_{\ell^1}))$$

$\rightsquigarrow$ Sparsity alone is not a good proxy for the sampling rate of ℓ^1-synthesis

Robustness to noise



(ix) If $m \gtrsim w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}_{\ell^1}))$ signal recovery via $(\mathrm{BP}^{\mathrm{sig}}_{\eta=0})$ is robust to measurement noise.

(x) If $\lambda_{\min}(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})) \ll 1$, coefficient recovery is less robust than signal recovery.

However, if $\lambda_{\min}(\boldsymbol{D}; \mathcal{D}_\wedge(\|\cdot\|_1, \boldsymbol{z}_{\ell^1})) \gg 1$, the contrary holds true.

Conclusion

- Coefficient/signal recovery via ℓ^1-synthesis with **Gaussian** measurements
- The sample complexities driven by $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}_{\ell^1}))$ lead to phase transitions
- Tight geometric upper-bound of $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}_{\ell^1}))$
- Illustration by extensive numerical XP
  - Sparsity alone is not a good proxy for the sampling rate of ℓ^1-synthesis
  - Need of a non-uniform theory across the class of all s-sparse signals
  - Robustness may differ between the recovered signal and coefficient

Conclusion

- Coefficient/signal recovery via ℓ^1-synthesis with **Gaussian** measurements
- The sample complexities driven by $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}_{\ell^1}))$ lead to phase transitions
- Tight geometric upper-bound of $w_\wedge^2(\boldsymbol{D} \cdot \mathcal{D}(\|\cdot\|_1\,; \boldsymbol{z}_{\ell^1}))$
- Illustration by extensive numerical XP
  - Sparsity alone is not a good proxy for the sampling rate of ℓ^1-synthesis
  - Need of a non-uniform theory across the class of all s-sparse signals
  - Robustness may differ between the recovered signal and coefficient

Thank you!