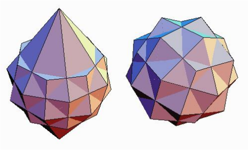


Sampling rates for ℓ^1-synthesis

C. Boyer, J. Kahn, M. März, P. Weiss

Séminaire Français d'optimisation, 03/12/2020

Context and objectives

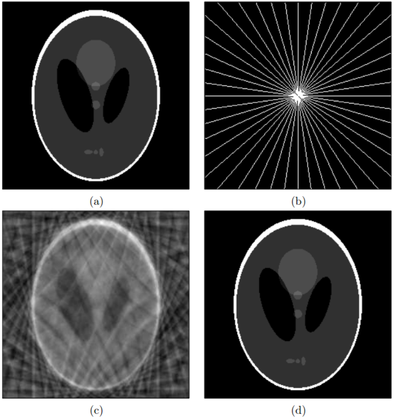


Figure: WHERE IT ALL STARTED: A PUZZLING EXPERIMENT...

Emmanuel J Candès, Justin Romberg, and Terence Tao.
Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information.
IEEE Transactions on information theory, 52(2):489–509, 2006.

Context and objectives

The problem

The image can be viewed as a vector $x_0 \in \mathbb{R}^n$.
The measurement operator is $A \in \mathbb{R}^{m \times n}$ (Fourier sub-sampling).
We observe $y = Ax_0$.

How to recover x_0 from y despite $m \ll n$?

The previous experiment

The estimate $\hat{x}$ was set as

$$\hat{x} \in \underset{x \in \mathbb{R}^n, Ax=y}{\operatorname{argmin}} \|\nabla x\|_1, \tag{1}$$

where $\nabla : \mathbb{R}^n \to \mathbb{R}^{2n}$ is a discrete version of the gradient operator.

Context and objectives

The typical accompanying theorems

Assume that

- The image x_0 is s-sparse with $m \geq Cs\log(n/s)$.
- The coefficients of $A \in \mathbb{R}^{m\times n}$ are i.i.d. normal random variables.
- Set

$$\hat{x} \in \underset{Ax=y}{\operatorname{argmin}} \|x\|_1. \tag{2}$$

Then $\hat{x} = x_0$ with high probability!

Emmanuel J Candès, Justin Romberg, and Terence Tao.
Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information.
IEEE Transactions on information theory, 52(2):489–509, 2006.

Context and objectives

The typical accompanying theorems

Assume that

- The image x_0 is s-sparse with $m \geq Cs\log(n/s)$.
- The coefficients of $A \in \mathbb{R}^{m\times n}$ are i.i.d. normal random variables.
- Set

$$\hat{x} \in \underset{Ax=y}{\operatorname{argmin}} \|x\|_1. \tag{2}$$

Then $\hat{x} = x_0$ with high probability!

Emmanuel J Candès, Justin Romberg, and Terence Tao.
Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information.
IEEE Transactions on information theory, 52(2):489–509, 2006.

A complete fraud?

- The matrix A was deterministic.
- We don't solve (2).
- The image has nothing sparse!

Context and objectives

The typical accompanying theorems

Assume that

- The image x_0 is s-sparse with $m \geq Cs\log(n/s)$.
- The coefficients of $A \in \mathbb{R}^{m\times n}$ are i.i.d. normal random variables.
- Set

$$\hat{x} \in \underset{Ax=y}{\operatorname{argmin}} \|x\|_1. \tag{2}$$

Then $\hat{x} = x_0$ with high probability!

Emmanuel J Candès, Justin Romberg, and Terence Tao.
Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information.
IEEE Transactions on information theory, 52(2):489–509, 2006.

A complete fraud?

- The matrix A was deterministic.
- We don't solve (2).
- The image has nothing sparse!

To date, the previous phenomenon is still not understood!

Context and objectives

This talk

- Compressed sensing is a success story.
- Now validated in various real life applications.
- Many facets in sampling theory are still obscure.

Objective: progress in understanding the previous experiment (and many others).

Context and objectives

This talk

- Compressed sensing is a success story.
- Now validated in various real life applications.
- Many facets in sampling theory are still obscure.

Objective: progress in understanding the previous experiment (and many others).

Disclaimer: we will fail.

Preliminaries

A more realistic (noisy) setting

We assume that:

$$y = Ax_0 + e$$

with $\|e\|_2 \leq \eta$ for $\eta \geq 0$.

The synthesis formulation (this talk)

We consider a dictionary (possibly $d \gg n$):

$$D = [d_1, \ldots, d_d] \in \mathbb{R}^{n \times d}$$

The synthesis formulation is:

$$\min_{z \in \mathbb{R}^d} \|z\|_1 \quad \text{s.t.} \quad \|ADz - y\|_2 \leq \eta \tag{$\mathcal{P}_z^\eta$}$$

We let $\hat{Z}$ denote the set of minimizers and $\hat{X} = D\hat{Z}$.

Preliminaries

The analysis formulation

Consider a linear transform $L = [l_1, \ldots, l_p] \in \mathbb{R}^{n \times p}$ (possibly $p \gg n$):
The analysis formulation is:

$$\min_{x \in \mathbb{R}^n} \|L^* x\|_1 \quad \text{s.t.} \quad \|Ax - y\|_2 \leq \eta. \qquad (\mathcal{A}_x^\eta)$$

Preliminaries

The analysis formulation

Consider a linear transform $L = [l_1, \dots, l_p] \in \mathbb{R}^{n \times p}$ (possibly $p \gg n$):
The analysis formulation is:

$$\min_{x \in \mathbb{R}^n} \|L^* x\|_1 \quad \text{s.t.} \quad \|Ax - y\|_2 \leq \eta. \tag{$\mathcal{A}_x^\eta$}$$

Theorem 1 (Analysis VS Synthesis).

Let $g_K(x) := \inf_{\lambda > 0} \{x \in \lambda K\}$ *denote the gauge of a convex* K*.*
Let $g_D := g_{DB_1^d}$*, then we have*

$$\hat{X} = \inf_{x \in \mathbb{R}^n} g_D(x) \quad s.t. \quad \|Ax - y\|_2 \leq \eta. \tag{$\mathcal{P}_x^\eta$}$$

If $\{d_1, \dots, d_d\} = \mathrm{Ext}(\{x, \|L^* x\|_1 \leq 1\})$*:*

$$\mathcal{P}_x^\eta = \mathcal{A}_x^\eta$$

Preliminaries

The analysis formulation

Consider a linear transform $L = [l_1, \ldots, l_p] \in \mathbb{R}^{n \times p}$ (possibly $p \gg n$):
The analysis formulation is:

$$\min_{x \in \mathbb{R}^n} \|L^* x\|_1 \quad \text{s.t.} \quad \|Ax - y\|_2 \leq \eta. \qquad (\mathcal{A}_x^\eta)$$

Theorem 1 (Analysis VS Synthesis).

Let $g_K(x) := \inf_{\lambda > 0} \{x \in \lambda K\}$ *denote the gauge of a convex* K*.*
Let $g_D := g_{DB_1^d}$*, then we have*

$$\hat{X} = \inf_{x \in \mathbb{R}^n} g_D(x) \quad s.t. \quad \|Ax - y\|_2 \leq \eta. \qquad (\mathcal{P}_x^\eta)$$

If $\{d_1, \ldots, d_d\} = \mathrm{Ext}(\{x, \|L^* x\|_1 \leq 1\})$*:*

$$\mathcal{P}_x^\eta = \mathcal{A}_x^\eta$$

Analysis problems $\subsetneq$ Synthesis problems

But: combinatorial explosion of extreme points for the analysis.

Preliminaries

A generic regularization framework

Let $f : \mathbb{R}^n \to \mathbb{R}$ denote a convex function.

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{s.t.} \qquad \|Ax - y\|_2 \leq \eta,$$

Definition 2 (Descent cone).

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and $x_0 \in \mathbb{R}^n$.
The **descent set** of f at x_0 is given by

$$\mathcal{D}_{f,x_0} := \{h \in \mathbb{R}^n : f(x_0 + h) \leq f(x_0)\}, \tag{3}$$

The **descent cone** is defined by $\mathcal{D}_\wedge(f, x_0) := \text{cone}(\mathcal{D}_{f,x_0})$.

Definition 3 (Minimum conic singular value).

Consider $A \in \mathbb{R}^{m \times n}$ and a cone $K \subseteq \mathbb{R}^n$.
The minimum conic singular value of A relative to K is:

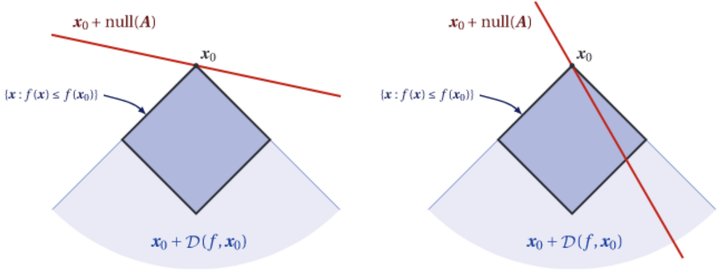
$$\lambda_{\min}(A, K) := \inf_{x \in K \cap \mathbb{S}^{n-1}} \|Ax\|_2. \tag{4}$$

Preliminaries

Theorem 4 (Recovery conditions).

- *If $\eta = 0$ then the following are equivalent:*
  - $\hat{x} = x_0$
  - $\ker(A) \cap \mathcal{D}_\wedge(f, x_0) = \{0\}$
  - $\lambda_{\min}(A, \mathcal{D}_\wedge(f, x_0)) > 0$
- *For $\eta > 0$, any solution $\hat{x}$ of the previous convex program satisfies:*

$$\|\hat{x} - x_0\|_2 \leq \frac{2\eta}{\lambda_{\min}(A, \mathcal{D}_\wedge(f, x_0))}$$



Dennis Amelunxen, Martin Lotz, Michael B McCoy, and Joel A Tropp.
Living on the edge: Phase transitions in convex programs with random data.
Information and Inference: A Journal of the IMA, 3(3):224–294, 2014.

Preliminaries

A useful condition?

For arbitrary A and convex cone K

$$\text{Testing } \lambda_{\min}(A, K) > 0 \text{ is NP-hard!}$$

(This is an instance of testing the co-positivity of matrices)

Katta G Murty and Santosh N Kabadi.
Some np-complete problems in quadratic and nonlinear programming.
Technical report, 1985.

Preliminaries

Definition 5 (Mean width).

- The mean width of a set $K \in \mathbb{R}^n$ is

$$w(K) = \mathbb{E}\left(\sup_{h \in K} \langle g, h\rangle\right) \text{ with } g \sim \mathcal{N}(0, I_n).$$

- The conic mean width of a cone $K \in \mathbb{R}^n$ is

$$w_\wedge(K) = w(K \cap \mathbb{S}^{n-1})$$

Theorem 6 (Generic recovery).

Assume that $A \in \mathbb{R}^{m \times n}$ is a Gaussian random matrix then

$$\lambda_{\min}(A, K) \geq \sqrt{m-1} - w_\wedge(K) - u$$

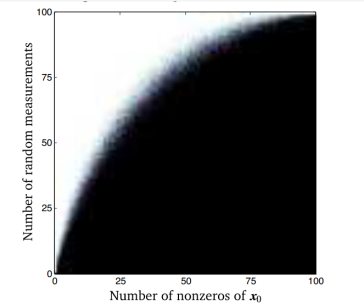
with probability larger than $1 - e^{-u^2/2}$.
A sufficient condition for robust recovery is

$$m \geq w_\wedge(\mathcal{D}_\wedge(f, x_0))^2 + 1.$$

Preliminaries

Theorem 7 (Phase transitions).

- *For* $m \geq w_\wedge(\mathcal{D}_\wedge(f, x_0))^2 + 1 - \log(\epsilon)\sqrt{n}$ *succeeds with probability* $> 1 - \epsilon$.
- *For* $m \leq w_\wedge(\mathcal{D}_\wedge(f, x_0))^2 + 1 + \log(\epsilon)\sqrt{n}$ *succeeds with probability* $< \epsilon$.



Dennis Amelunxen, Martin Lotz, Michael B McCoy, and Joel A Tropp.
Living on the edge: Phase transitions in convex programs with random data.
Information and Inference: A Journal of the IMA, 3(3):224–294, 2014.

Sampling rates

Definition 8 (Minimal ℓ^1 representers).

Let $x_0 \in \mathbb{R}^n$ denote a signal and $D \in \mathbb{R}^{n \times d}$ be a dictionary.
The set of minimal ℓ^1 representers of x_0 w.r.t. D is

$$Z_{\ell^1} := \underset{Dz = x_0}{\operatorname{argmin}} \|z\|_1.$$

Remark (Uniqueness).

*For a dictionary in **generic** position, $Z_{\ell^1} = \{z_{\ell^1}\}$ is a singleton.*
For practical dictionaries (e.g. redundant wavelets) it is not.

Sampling rates

Definition 8 (Minimal ℓ^1 representers).

Let $x_0 \in \mathbb{R}^n$ denote a signal and $D \in \mathbb{R}^{n\times d}$ be a dictionary.
The set of minimal ℓ^1 representers of x_0 w.r.t. D is

$$Z_{\ell^1} := \underset{Dz=x_0}{\operatorname{argmin}} \|z\|_1.$$

Remark (Uniqueness).

*For a dictionary in **generic** position, $Z_{\ell^1} = \{z_{\ell^1}\}$ is a singleton.*
For practical dictionaries (e.g. redundant wavelets) it is not.

Sampling rates

Remark (A necessary condition for coefficient recovery).

Depending on the application we may solve $\mathcal{P}_z^\eta$ or $\mathcal{P}_x^\eta$.
Recovering z_0 s.t. $x_0 = Dz_0$ is only possible if $Z_{\ell^1} = \{z_0\}$.

Proposition (Descent cone of the gauge).

Let x_0 denote a vector with minimal l^1 representers Z_{ℓ^1}. Then

$$\mathcal{D}_\wedge(g_D, x_0) = D \cdot \mathcal{D}_\wedge(\| \cdot \|_1, z_{\ell^1}) \quad \text{for all} \quad z_{\ell^1} \in Z_{\ell^1}.$$

Sampling rates

Theorem 9 (Sampling rates for coefficients (noisy)).

Let $z_0 \in \mathbb{R}^d$ *be a coefficient and* $y = ADz_0 + e$. *Then if*

$$Z_{\ell^1} = \{z_0\} \quad \textit{and} \quad m \geq m_0 := (w_\wedge(D \cdot \mathcal{D}_\wedge(\|\cdot\|_1, z_0)) + u)^2 + 1,$$

any solution $\hat{z}$ *of* $(\mathcal{P}_z^\eta)$ *satisfies with probability* $> 1 - e^{-u^2/2}$

$$\|z_0 - \hat{z}\|_2 \leq \frac{2\eta}{\lambda_{\min}(D, \mathcal{D}_\wedge(\|\cdot\|_1, z_0))(\sqrt{m-1} - \sqrt{m_0 - 1})}.$$

Sampling rates

Theorem 9 (Sampling rates for coefficients (noisy)).

Let $z_0 \in \mathbb{R}^d$ *be a coefficient and* $y = ADz_0 + e$. *Then if*

$$Z_{\ell^1} = \{z_0\} \quad \text{and} \quad m \geq m_0 := (w_\wedge(D \cdot \mathcal{D}_\wedge(\|\cdot\|_1, z_0)) + u)^2 + 1,$$

any solution $\hat{z}$ *of* $(\mathcal{P}_z^\eta)$ *satisfies with probability* $> 1 - e^{-u^2/2}$

$$\|z_0 - \hat{z}\|_2 \leq \frac{2\eta}{\lambda_{\min}(D, \mathcal{D}_\wedge(\|\cdot\|_1, z_0))(\sqrt{m-1} - \sqrt{m_0 - 1})}.$$

Theorem 10 (Sampling rates for signal).

Let $x_0 \in \mathbb{R}^n$ *be a signal and* $y = Ax_0 + e$. *Then for* $m \geq m_0$ *any solution* $\hat{x}$ *of*

$$\min_{x \in \mathbb{R}^n} g_D(x) \quad s.t. \quad \|Ax - y\|_2 \leq \eta.$$

satisfies:

$$\|x_0 - \hat{x}\|_2 \leq \frac{2\eta}{\sqrt{m-1} - \sqrt{m_0 - 1}}.$$

Sampling rates

Conclusion

Sampling rates coefficients = sampling rate signal.

Stability to noise is different.

Critical quantity = conic mean width of a linearly transformed cone

$$w_\wedge(D \cdot \mathcal{D}_\wedge(\| \cdot \|_1, z_0)).$$

Sampling rates

Uniform recovery VS non uniform recovery

Compressed sensing started with the Restricted Isometry Property leading to:

All s-sparse vectors are recovered with probability X if $m > m_{RIP}$.

Here:

A specific vector z_0 is recovered with probability Y if $m > m_{z_0}$.

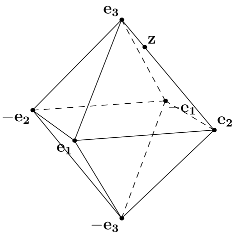
The Restricted Isometry Property...

- RIP = far stronger statement.
- RIP = optimal for orthogonal D, super pessimistic otherwise.
- RIP = useless in 99% of the practical cases.

Upper bounds on the conic Gaussian width

Understanding the descent cone

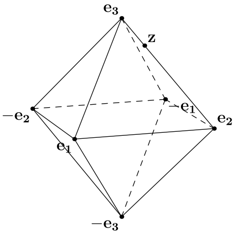
A cryptic quantity: $w_\wedge(D \cdot \mathcal{D}_\wedge(\|\cdot\|_1, z))$.

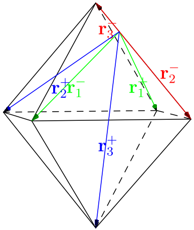


Upper bounds on the conic Gaussian width

Understanding the descent cone

A cryptic quantity: $w_\wedge(D \cdot \mathcal{D}_\wedge(\|\cdot\|_1, z))$.

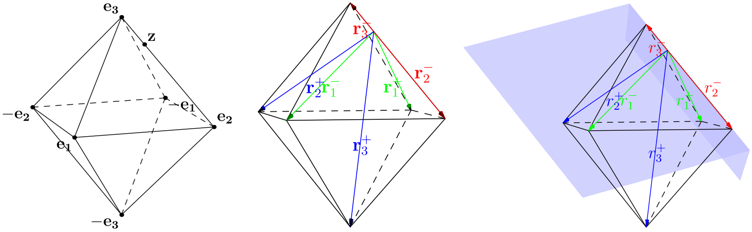




Upper bounds on the conic Gaussian width

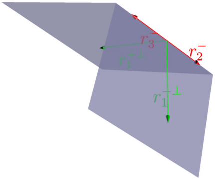
Understanding the descent cone

A cryptic quantity: $w_\wedge(D \cdot \mathcal{D}_\wedge(\|\cdot\|_1, z))$.

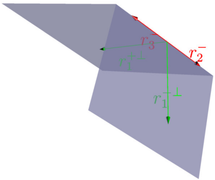


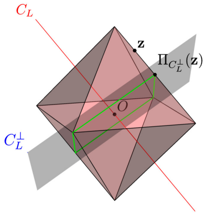
$$\mathcal{D}_\wedge(\|\cdot\|_1, z_0) = \mathrm{cone}(z \pm \|z\|_1 e_i, 1 \leq i \leq d)$$

Upper bounds on the conic Gaussian width

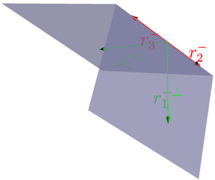


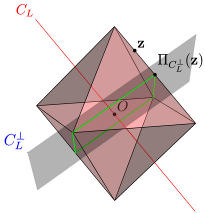
Upper bounds on the conic Gaussian width

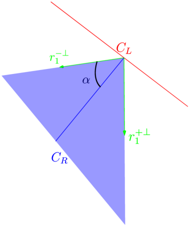




Upper bounds on the conic Gaussian width







Theorem 11 (Orthogonal decomposition).

$$C = D\mathcal{D}_\wedge(\|\cdot\|_1, z_0) = C_L \oplus C_R$$

with for any $z \in \mathrm{ri}(Z_{\ell^1})$, $\bar{s} = \|z\|_0$*:*

$$C_L = \mathrm{span}(x_0 - \bar{s} \cdot \mathrm{sign}(z_i) d_i)$$

$$C_R = \mathrm{cone}(r_i^\pm, i \in S^c) \text{ with } r_i^\pm := \Pi_{C_L^\perp}(Dz \pm \bar{s} \cdot \mathrm{sign}(z_i) d_i)$$

Upper bounds on the conic Gaussian width

Definition 12 (Circumangle).

Let $C \subsetneq \mathbb{R}^n$ denote a closed convex cone.
The **circumangle** $\alpha(C)$ is defined by:

$$\cos(\alpha) = \sup_{\|\theta\|_2=1} \inf_{x \in C, \|x\|_2=1} \langle x, \theta \rangle$$

Upper bounds on the conic Gaussian width

Definition 12 (Circumangle).

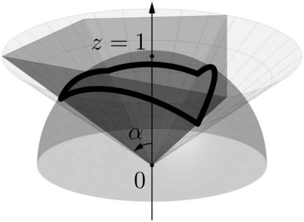
Let $C \subsetneq \mathbb{R}^n$ denote a closed convex cone.
The **circumangle** $\alpha(C)$ is defined by:

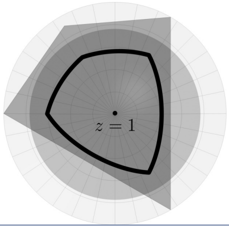
$$\cos(\alpha) = \sup_{\|\theta\|_2=1} \inf_{x \in C, \|x\|_2=1} \langle x, \theta \rangle$$

Theorem 13 (Width of polyhedral α cones).

Let C_k^α denote the set of k-polyhedral α cones in $\mathbb{R}^n$.
Then

$$\sup_{C \in C_k^\alpha} w_\wedge(C) = O(\tan(\alpha)\sqrt{2\log(k)}).$$





Upper bounds on the conic Gaussian width

Theorem 14 (An upper bound on the sampling rate).

Let $\alpha =$ *circumangle of* C_R. *We have:*

$$w_\wedge^2(\mathcal{D}_\wedge(g_D, x_0)) \leq \bar{s} + C\tan(\alpha)^2\log(d)$$

The critical number of measurements is

$$m_0 \equiv \bar{s} + \tan(\alpha)^2\log(d).$$

Computing $\alpha \equiv$ *convex problem!*

Upper bounds on the conic Gaussian width

Theorem 14 (An upper bound on the sampling rate).

Let $\alpha =$ *circumangle of* C_R*. We have:*

$$w_\wedge^2(\mathcal{D}_\wedge(g_D, x_0)) \leq \bar{s} + C \tan(\alpha)^2 \log(d)$$

The critical number of measurements is

$$m_0 \equiv \bar{s} + \tan(\alpha)^2 \log(d).$$

Computing $\alpha \equiv$ *convex problem!*

Ongoing...

Show that the previous bound:

- recovers all the best existing results.
- allows to establish new ones (super-resolution, wavelet frames).

What about total gradient variation?

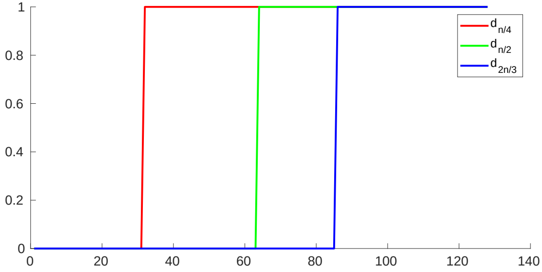
Example with 1D total variation

Let $y = Ax_0 + e$, with $\bar{x}_0 = 0$. Consider:

$$\min_{x \in \mathbb{R}^n} \|Lx\|_1 \text{ s.t. } \|Ax - y\|_2 \leq \eta \tag{TV_{1D}}$$

where $(Lx)_i = x_{i+1} - x_i$. Then (TV_{1D}) is equivalent to

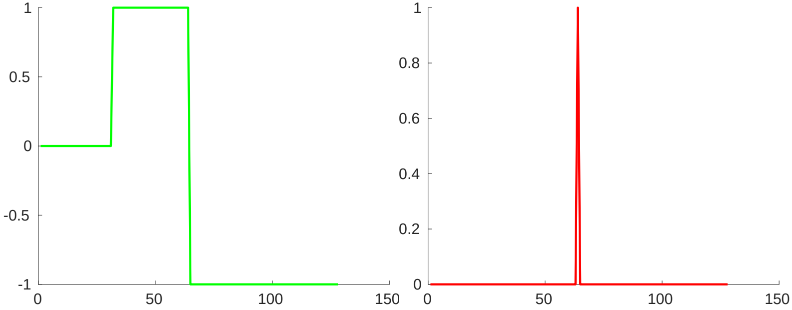
$$\min_{z \in \mathbb{R}^{n-1}} \|z\|_1 \text{ s.t. } \|ADz - y\|_2 \leq \eta \quad \text{with} \quad D = [d_1, \dots, d_{n-1}]$$



What about total gradient variation?

Early remarks

- $\text{Coherence}(D) = \max_{i \neq j} \dfrac{|\langle d_i, d_j \rangle|}{\|d_i\|_2 \cdot \|d_j\|_2} \simeq 1.$
- **Usual theories cannot work**: sparsity is not the right complexity measure!



(a) $m_0 \simeq 5$. (b) m_0 very large!

Figure: Different 2-sparse vectors have very different complexities.

What about total gradient variation?

Theorem 15 (1D TV and the circumangle).

x_0: piecewise constant with s jumps with signs (σ_k) at $(p_k)_{1\le k\le s}$.

Then

$$m_0 \le s + C\sqrt{\frac{n}{4L}}\log(n)^{3/2}$$

With

$$L = \frac{1}{\sum_k |(\sigma_{k+1}-\sigma_k)|/(p_{k-1}-p_k)}$$

$L \equiv$ *harmonic mean of the distances between jumps of opposite signs.*

Similar conclusions in:

Francesco Ortelli and Sara van de Geer.
Oracle inequalities for image denoising with total variation regularization.
Information&Inference, 2019.

Conclusion

Contributions

- ✔ Sampling rates for the synthesis problem (coefficient and signal).
- ✔ Decent upper-bounds for the conic width of linearly transformed cones.
- ✔ Dissected the descent cone of the ℓ^1-ball.
- × Quantities are still partly cryptic.
- × Case by case study of practical dictionaries is technical.

Dennis Amelunxen, Martin Lotz, Michael B McCoy, and Joel A Tropp.
Living on the edge: Phase transitions in convex programs with random data.
Information and Inference: A Journal of the IMA, 3(3):224–294, 2014.

Claire Boyer, Jonas Kahn, Maximilian März, and Pierre Weiss.
Sampling rates for ℓ^1-synthesis.
preprint, 2020.

Thanks for your attention