# Signal Recovery, Uncertainty Relations, and Minkowski Dimension

Helmut Bőlcskei

ETH Zurich

December 2013

Joint work with C. Aubel, P. Kuppinger, G. Pope, E. Riegler, D. Stotz, and C. Studer

# Aim of this Talk

- Develop a unified theory for a wide range of (sparse) signal recovery problems:

# Aim of this Talk

- Develop a unified theory for a wide range of (sparse) signal recovery problems:
  - Signal separation

# Aim of this Talk

- Develop a unified theory for a wide range of (sparse) signal recovery problems:
    - Signal separation
    - Super-resolution

# Aim of this Talk

- Develop a unified theory for a wide range of (sparse) signal recovery problems:
  - Signal separation
  - Super-resolution
  - Inpainting

# Aim of this Talk

- Develop a unified theory for a wide range of (sparse) signal recovery problems:
  - Signal separation
  - Super-resolution
  - Inpainting
  - De-clipping

# Aim of this Talk

- Develop a unified theory for a wide range of (sparse) signal recovery problems:
  - Signal separation
  - Super-resolution
  - Inpainting
  - De-clipping
  - Removal of impulse noise or narrowband interference

# Aim of this Talk

- Develop a unified theory for a wide range of (sparse) signal recovery problems:
  - Signal separation
  - Super-resolution
  - Inpainting
  - De-clipping
  - Removal of impulse noise or narrowband interference
- Establish fundamental performance limits

# Aim of this Talk

- Develop a unified theory for a wide range of (sparse) signal recovery problems:
  - Signal separation
  - Super-resolution
  - Inpainting
  - De-clipping
  - Removal of impulse noise or narrowband interference
- Establish fundamental performance limits
- Propose an information-theoretic formulation

# Signal Separation

Decompose image into cartoon and textured part

observation

# Signal Separation

Decompose image into cartoon and textured part

observation

cartoon

texture

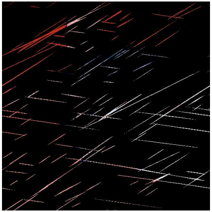
# Image Restoration

observation

# Image Restoration

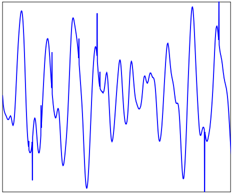
observation

signal



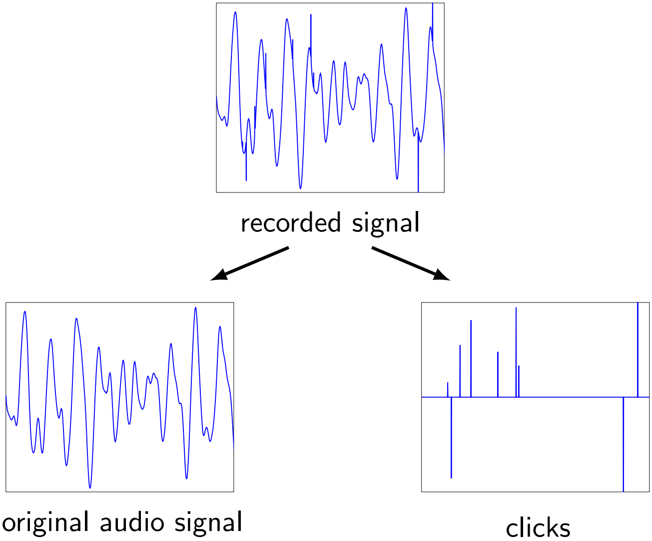
scratches

# Removing “Clicks” from a Vinyl/Record Player



recorded signal

# Removing "Clicks" from a Vinyl/Record Player



# Structural Specifics and Signal Model

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$$

- Transform $\mathbf{A}$ "sparsifies" images, e.g., wavelet transform

# Structural Specifics and Signal Model

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$$

- Transform $\mathbf{A}$ "sparsifies" images, e.g., wavelet transform
- "Error" signal sparse in transform $\mathbf{B}$:

# Structural Specifics and Signal Model

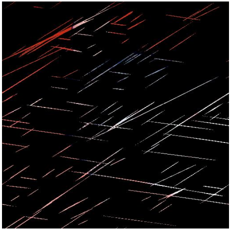
$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$$

- Transform $\mathbf{A}$ "sparsifies" images, e.g., wavelet transform
- "Error" signal sparse in transform $\mathbf{B}$:
  - texture: sparse in curvelet frame

# Structural Specifics and Signal Model

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$$

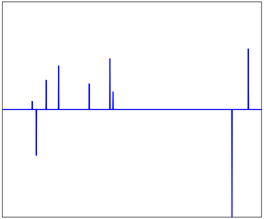
- Transform $\mathbf{A}$ "sparsifies" images, e.g., wavelet transform
- "Error" signal sparse in transform $\mathbf{B}$:
  - texture: sparse in curvelet frame
  - scratches: sparse in ridgelet frame

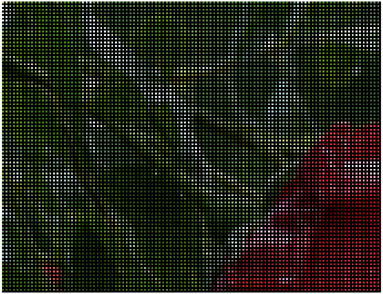


# Structural Specifics and Signal Model

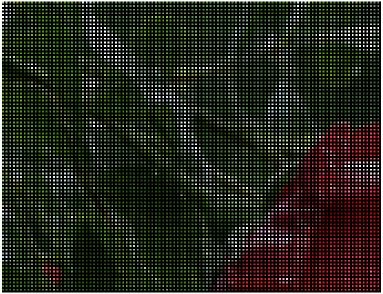
$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$$

- Transform $\mathbf{A}$ "sparsifies" images, e.g., wavelet transform
- "Error" signal sparse in transform $\mathbf{B}$:
  - texture: sparse in curvelet frame
  - scratches: sparse in ridgelet frame
  - clicks: sparse in identity basis





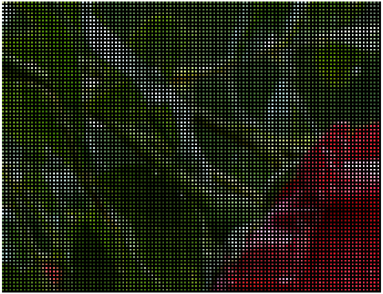
Downsampled image
(by a factor of 9)



Downsampled image
(by a factor of 9)

Linear interpolation

# Super-Resolution



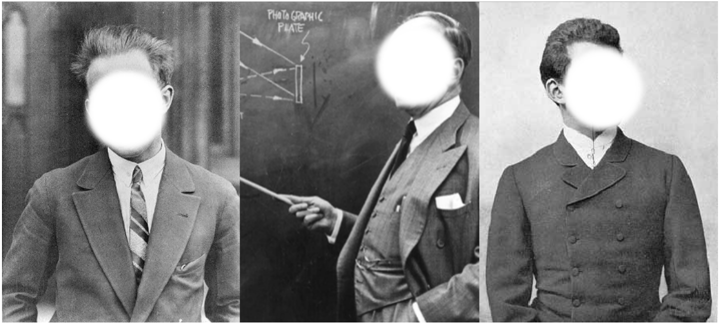
Downsampled image

(by a factor of 9)

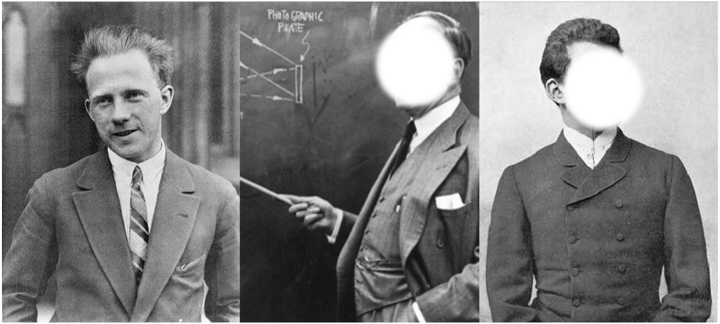
Sparsity-exploiting

reconstruction

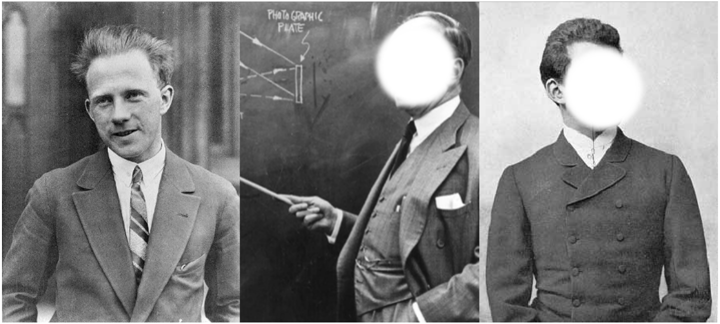
# Inpainting



# Inpainting

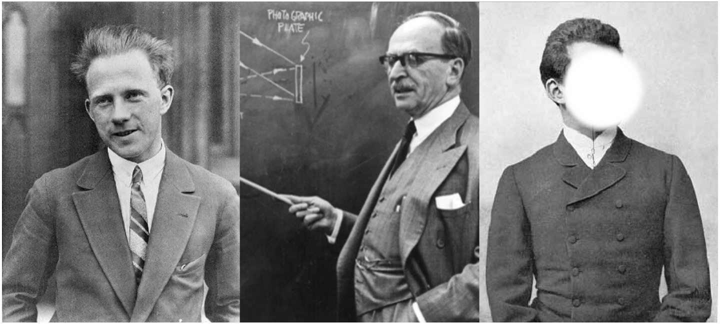


# Inpainting



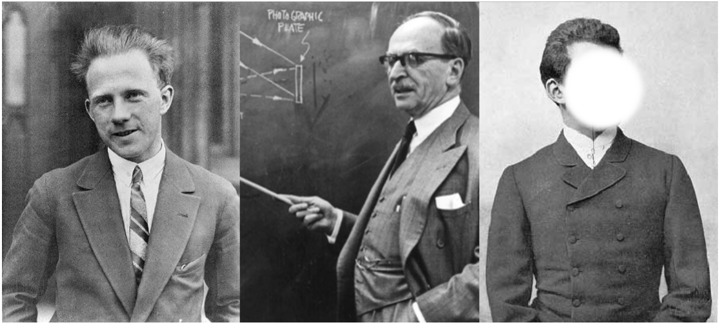

W. Heisenberg

# Inpainting



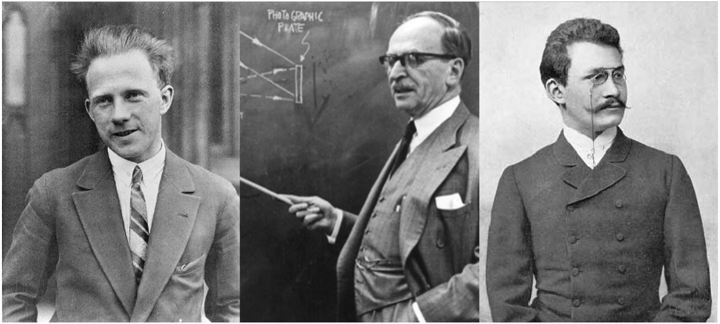

W. Heisenberg

# Inpainting



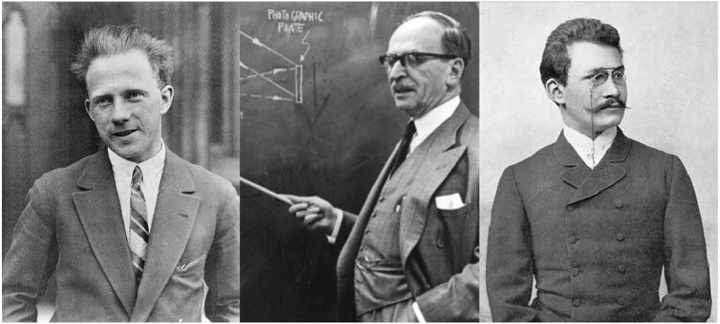

W. Heisenberg D. Gábor

# Inpainting




W. Heisenberg D. Gábor

# Inpainting




W. Heisenberg D. Gábor H. Minkowski

# Signal Model for Super-Resolution and Inpainting

- Only a subset of the entries in

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

  is available

# Signal Model for Super-Resolution and Inpainting

- Only a subset of the entries in

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

  is available

- Taken into account by assuming that we observe

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} = \mathbf{A}\mathbf{x} + \mathbf{I}\mathbf{e}$$

  and choosing $\mathbf{e}$ so that the missing entries of $\mathbf{y}$ are set to, e.g., $0$

# Signal Model for Super-Resolution and Inpainting

- Only a subset of the entries in

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

  is available

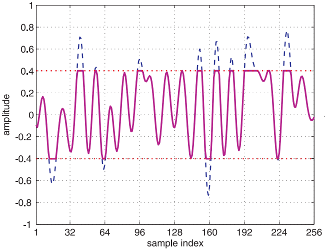
- Taken into account by assuming that we observe

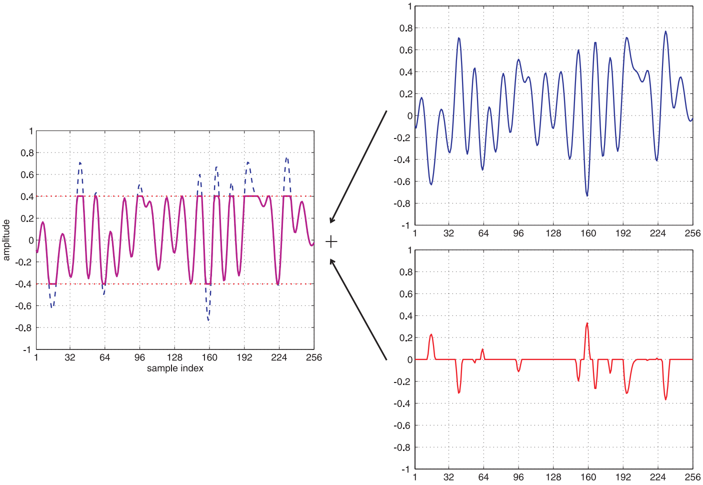
$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} = \mathbf{A}\mathbf{x} + \mathbf{I}\mathbf{e}$$

  and choosing $\mathbf{e}$ so that the missing entries of $\mathbf{y}$ are set to, e.g., $0$

- "Error" signal $\mathbf{e}$ is sparse if few entries are missing

# Recovery of Clipped Signals




1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
1
32
64
96
128
160
192
224
256
sample index
amplitude
+

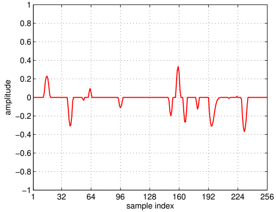
# Signal Model for Clipping

- Instead of

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

we observe

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \underbrace{[\mathrm{clip}(\mathbf{A}\mathbf{x}) - \mathbf{A}\mathbf{x}]}_{\text{sparse in } \mathbf{B}=\mathbf{I}}$$



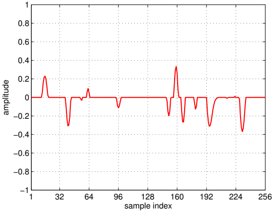
# Signal Model for Clipping

- Instead of

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

we observe

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \underbrace{[\mathrm{clip}(\mathbf{A}\mathbf{x}) - \mathbf{A}\mathbf{x}]}_{\text{sparse in } \mathbf{B}=\mathbf{I}}$$




- "Error" signal is sparse if clipping is not too aggressive

- Support set of $\mathbf{e}$ is known

# Some Existing Approaches

Literature is rich, e.g.

- Signal separation:
  - Morphological component analysis [*Starck et al., 2004; Elad et al., 2005*]
  - Split-Bregman methods [*Cai et al., 2009*]
  - Microlocal analysis [*Donoho & Kutyniok, 2010*]
  - Convex demixing [*McCoy & Tropp, 2013*]

# Some Existing Approaches

Literature is rich, e.g.

- Signal separation:
  - Morphological component analysis [*Starck et al., 2004; Elad et al., 2005*]
  - Split-Bregman methods [*Cai et al., 2009*]
  - Microlocal analysis [*Donoho & Kutyniok, 2010*]
  - Convex demixing [*McCoy & Tropp, 2013*]
- Super-resolution:
  - Navier-Stokes [*Bertalmio et al., 2001*]
  - Sparsity enhancing [*Yang et al., 2008*]
  - Total variation minimization [*Candès & Fernandez-Granda, 2013*]

# Some Existing Approaches Cont'd

- Inpainting:
    - Local transforms and separation [*Dong et al., 2011*]
    - Total variation minimization [*Chambolle, 2004*]
    - Morphological component analysis [*Elad et al., 2005*]
    - Image colorization [*Sapiro, 2005*]
    - Clustered sparsity [*King et al., 2014*]
- De-clipping:
    - Constrained matching pursuit [*Adler et al., 2011*]

# General Problem Statement

- Signal model:

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$$

# General Problem Statement

- Signal model:

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$$

  - $\mathbf{x}$, $\mathbf{e}$ sparse, may depend on each other

# General Problem Statement

- Signal model:

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$$

  - $\mathbf{x}$, $\mathbf{e}$ sparse, may depend on each other
  - $\mathbf{A}$, $\mathbf{B}$ dictionaries (bases, incomplete sets, or frames)

# General Problem Statement

- Signal model:

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$$

  - $\mathbf{x}$, $\mathbf{e}$ sparse, may depend on each other
  - $\mathbf{A}$, $\mathbf{B}$ dictionaries (bases, incomplete sets, or frames)
  - Redundancy can lead to sparser representation

# General Problem Statement

- Signal model:

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$$

  - $\mathbf{x}$, $\mathbf{e}$ sparse, may depend on each other
  - $\mathbf{A}$, $\mathbf{B}$ dictionaries (bases, incomplete sets, or frames)
  - Redundancy can lead to sparser representation

Examples:

- Overcomplete DFT
- Gabor frames
- Curvelet or wavelet frames
- Ridgelets or shearlets

## General Problem Statement

- Signal model:

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$$

  - $\mathbf{x}$, $\mathbf{e}$ sparse, may depend on each other
  - $\mathbf{A}$, $\mathbf{B}$ dictionaries (bases, incomplete sets, or frames)
  - Redundancy can lead to sparser representation

Examples:

- Overcomplete DFT
- Gabor frames
- Curvelet or wavelet frames
- Ridgelets or shearlets

Want to recover $\mathbf{x}$ and/or $\mathbf{e}$ from $\mathbf{z}$!
Knowledge on $\mathbf{x}$ and/or $\mathbf{e}$ may be available
(support set, sparsity level, full knowledge).

# Formalizing the Problem

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} = \underbrace{[\mathbf{A} \quad \mathbf{B}]}_{\mathbf{D}} \begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix}$$

# Formalizing the Problem

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} = \underbrace{[\mathbf{A} \quad \mathbf{B}]}_{\mathbf{D}} \begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix}$$

- Requires solving an **underdetermined linear system of equations**

# Formalizing the Problem

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} = \underbrace{[\mathbf{A} \;\; \mathbf{B}]}_{\mathbf{D}} \begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix}$$

- Requires solving an **underdetermined linear system of equations**
- What are the **fundamental limits** on extracting $\mathbf{x}$ and $\mathbf{e}$ from $\mathbf{z}$?

# Formalizing the Problem

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} = \underbrace{[\mathbf{A} \;\; \mathbf{B}]}_{\mathbf{D}} \begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix}$$

- Requires solving an **underdetermined linear system of equations**
- What are the **fundamental limits** on extracting $\mathbf{x}$ and $\mathbf{e}$ from $\mathbf{z}$?
- Could use $\frac{1}{2}(1 + 1/\mu)$-threshold [*Donoho & Elad, 2003; Gribonval & Nielsen, 2003*] for general $\mathbf{D}$

# Uniqueness

- Assume there exist two pairs $(\mathbf{x}, \mathbf{e})$ and $(\mathbf{x}', \mathbf{e}')$ such that

$$\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} = \mathbf{A}\mathbf{x}' + \mathbf{B}\mathbf{e}'$$

  and hence

$$\mathbf{A}(\mathbf{x} - \mathbf{x}') = \mathbf{B}(\mathbf{e}' - \mathbf{e})$$

# Uniqueness

- Assume there exist two pairs $(\mathbf{x}, \mathbf{e})$ and $(\mathbf{x}', \mathbf{e}')$ such that

$$\mathbf{Ax} + \mathbf{Be} = \mathbf{Ax}' + \mathbf{Be}'$$

  and hence

$$\mathbf{A}(\mathbf{x} - \mathbf{x}') = \mathbf{B}(\mathbf{e}' - \mathbf{e})$$

- The vectors $(\mathbf{x} - \mathbf{x}')$ and $(\mathbf{e}' - \mathbf{e})$ represent the same signal $\mathbf{s}$

$$\mathbf{A}(\mathbf{x} - \mathbf{x}') = \mathbf{B}(\mathbf{e}' - \mathbf{e}) \triangleq \mathbf{s}$$

  in two different dictionaries $\mathbf{A}$ and $\mathbf{B}$

# Enter Uncertainty Principle

- Assume that
  - $\mathbf{x}$, $\mathbf{x}'$ are $n_x$-sparse $\quad \Rightarrow \quad \mathbf{x} - \mathbf{x}'$ is $(2n_x)$-sparse
  - $\mathbf{e}$, $\mathbf{e}'$ are $n_e$-sparse $\quad \Rightarrow \quad \mathbf{e}' - \mathbf{e}$ is $(2n_e)$-sparse

# Enter Uncertainty Principle

- Assume that
  - $\mathbf{x}$, $\mathbf{x}'$ are $n_x$-sparse $\quad\Rightarrow\quad$ $\mathbf{x}-\mathbf{x}'$ is $(2n_x)$-sparse
  - $\mathbf{e}$, $\mathbf{e}'$ are $n_e$-sparse $\quad\Rightarrow\quad$ $\mathbf{e}'-\mathbf{e}$ is $(2n_e)$-sparse
- If
  - $n_x$ and $n_e$ are "small enough"
  - $\mathbf{A}$ and $\mathbf{B}$ are "sufficiently different"

  it may not be possible to satisfy

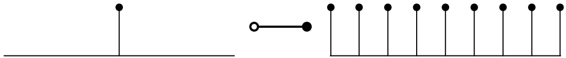
$$\mathbf{s} = \mathbf{A}(\mathbf{x}-\mathbf{x}') = \mathbf{B}(\mathbf{e}'-\mathbf{e})$$

# Uncertainty Relations for ONBs

- [*Donoho & Stark, 1989*]: $\mathbf{A} = \mathbf{I}_m$, $\mathbf{B} = \mathbf{F}_m$, $\mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q}$, then

$$\|\mathbf{p}\|_0 \, \|\mathbf{q}\|_0 \geqslant m$$

$m$-point DFT



- [*Elad & Bruckstein, 2002*]: $\mathbf{A}$ and $\mathbf{B}$ general ONBs with $\mu \triangleq \max_{i \neq j} |\langle \mathbf{a}_i, \mathbf{b}_j \rangle|$, then

$$\|\mathbf{p}\|_0 \, \|\mathbf{q}\|_0 \geqslant \frac{1}{\mu^2}$$

# Uncertainty Relation for General $\mathbf{A}$, $\mathbf{B}$

**Theorem (Studer et al., 2011)**

*Let*

- $\mathbf{A} \in \mathbb{C}^{m \times n_a}$ *be a dictionary with coherence* $\mu_a$
- $\mathbf{B} \in \mathbb{C}^{m \times n_b}$ *be a dictionary with coherence* $\mu_b$
- $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$ *have coherence* $\mu$
- $\mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q}$

*Then, we have*

$$\|\mathbf{p}\|_0 \|\mathbf{q}\|_0 \geqslant \frac{[1 - \mu_a(\|\mathbf{p}\|_0 - 1)]^+ [1 - \mu_b(\|\mathbf{q}\|_0 - 1)]^+}{\mu^2}.$$

# Recovery with BP if $\mathrm{supp}(\mathbf{e})$ is Known (e.g., Declipping)

**Theorem (Studer et al., 2011)**

*Let* $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$ *where* $\mathcal{E} = supp(\mathbf{e})$ *is known. Consider the convex program*

$$(BP, \mathcal{E}) \quad \begin{cases} minimize & \|\tilde{\mathbf{x}}\|_1 \\ subject\ to & \mathbf{A}\tilde{\mathbf{x}} \in (\{\mathbf{z}\} + \mathcal{R}(\mathbf{B}_{\mathcal{E}})). \end{cases}$$

*If*

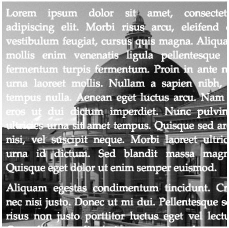
$$2\|\mathbf{x}\|_0 \, \|\mathbf{e}\|_0 < \frac{[1 - \mu_a(2\|\mathbf{x}\|_0 - 1)]^+ [1 - \mu_b(\|\mathbf{e}\|_0 - 1)]^+}{\mu^2}$$

*then the unique solution of* $(BP, \mathcal{E})$ *is given by* $\mathbf{x}$.

Extended to compressible signals and noisy measurements [*Studer & Baraniuk, 2011*]

**Example:** Separate text from picture

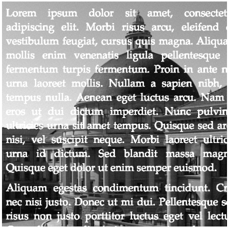
- Text is sparse in identity basis
- Use wavelets or DCT to sparsify image




Observation

**Example:** Separate text from picture

- Text is sparse in identity basis
- Use wavelets or DCT to sparsify image




Observation

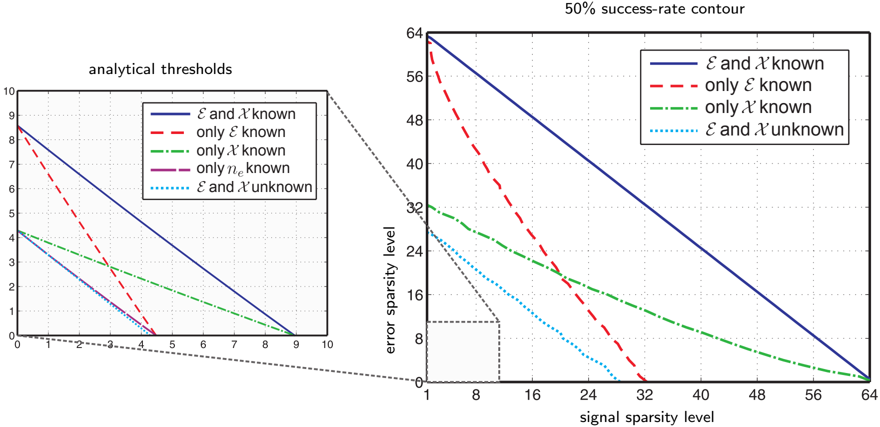
$\mathbf{A}$ = wavelet basis

$\mu = 0.25$

$\mathbf{A}$ = DCT

$\mu \approx 0.0039$

- Wavelet basis is more coherent with identity $\Rightarrow$ yields worse separation performance




- $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{64 \times 80}$
- $\mu_a \approx 0.126$, $\mu_b \approx 0.131$, and $\mu \approx 0.132$

# The Thresholds are Tight

$$\mathbf{A} = \mathbf{I}_m, \quad \mathbf{B} = \mathbf{F}_m, \quad \mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q}$$

# The Thresholds are Tight

$$\mathbf{A} = \mathbf{I}_m, \quad \mathbf{B} = \mathbf{F}_m, \quad \mathbf{Ap} = \mathbf{Bq}$$

- Recovery is unique if

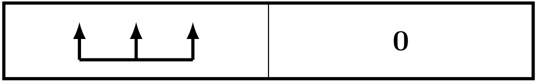
$$\|\mathbf{p}\|_0 + \|\mathbf{q}\|_0 < \sqrt{m}$$

# The Thresholds are Tight

$$\mathbf{A} = \mathbf{I}_m, \quad \mathbf{B} = \mathbf{F}_m, \quad \mathbf{Ap} = \mathbf{Bq}$$

- Recovery is unique if

$$\|\mathbf{p}\|_0 + \|\mathbf{q}\|_0 < \sqrt{m}$$



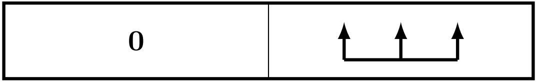

$$\begin{bmatrix} \delta & \mathbf{0} \end{bmatrix}$$

# The Thresholds are Tight

$$\mathbf{A} = \mathbf{I}_m, \quad \mathbf{B} = \mathbf{F}_m, \quad \mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q}$$

- Recovery is unique if

$$\|\mathbf{p}\|_0 + \|\mathbf{q}\|_0 < \sqrt{m}$$




$$\begin{bmatrix} \mathbf{0} & \boldsymbol{\delta} \end{bmatrix}$$

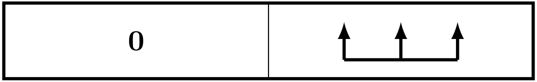
$$\begin{bmatrix} \mathbf{I} & \mathbf{F} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{F} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\delta} \end{bmatrix}$$

# The Thresholds are Tight

$$\mathbf{A} = \mathbf{I}_m, \quad \mathbf{B} = \mathbf{F}_m, \quad \mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q}$$

- Recovery is unique if

$$\|\mathbf{p}\|_0 + \|\mathbf{q}\|_0 < \sqrt{m}$$




$$\begin{bmatrix} \mathbf{0} & \boldsymbol{\delta} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{I} & \mathbf{F} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{F} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\delta} \end{bmatrix}$$

This behavior is fundamental and is known as the **square-root bottleneck**

# Probabilistic Recovery Guarantees for BP

- Neither support set known [*Kuppinger et al., 2011*]
- One or both support sets known [*Pope et al., 2011*]

Recovery possible with high probability even if

$$\|\mathbf{p}\|_0 + \|\mathbf{q}\|_0 \sim \frac{m}{\log n}$$

Compare to

$$\|\mathbf{p}\|_0 + \|\mathbf{q}\|_0 \sim \sqrt{m}$$

This "breaks" the **square-root bottleneck**!

# An Information-Theoretic Formulation

Sparsity for **random signals**:

$$\text{components of signal are drawn i.i.d.} \sim (1-\rho)\delta_0 + \rho P_{\mathsf{cont}}$$

where $0 \leqslant \rho \leqslant 1$ represents the mixture parameter

# An Information-Theoretic Formulation

Sparsity for **random signals**:

$$\text{components of signal are drawn i.i.d.} \sim (1-\rho)\delta_0 + \rho P_{\mathsf{cont}}$$

where $0 \leqslant \rho \leqslant 1$ represents the mixture parameter

$\longrightarrow$ For large dimensions, the fraction of nonzero components in the signal is given by $\rho$ (LLN)

# An Information-Theoretic Formulation

Sparsity for **random signals**:

components of signal are drawn i.i.d. $\sim (1-\rho)\delta_0 + \rho P_{\mathsf{cont}}$

where $0 \leqslant \rho \leqslant 1$ represents the mixture parameter

$\longrightarrow$ For large dimensions, the fraction of nonzero components in the signal is given by $\rho$ (LLN)

General distributions — Lebesgue decomposition:

$$P = \alpha P_{\mathsf{disc}} + \beta P_{\mathsf{cont}} + \gamma P_{\mathsf{sing}}, \quad \alpha + \beta + \gamma = 1$$

# Almost Lossless Signal Separation

Framework inspired by [*Wu & Verdú, 2010*]:

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$$

Existence of a measurable "separator" $g$ such that for general random sources **x**, **e**, for sufficiently large blocklengths

$$\mathbb{P}\left[g\left([\mathbf{A} \;\; \mathbf{B}]\begin{bmatrix}\mathbf{x}\\ \mathbf{e}\end{bmatrix}\right) \neq \begin{bmatrix}\mathbf{x}\\ \mathbf{e}\end{bmatrix}\right] < \varepsilon$$

# Almost Lossless Signal Separation

Framework inspired by [*Wu & Verdú, 2010*]:

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$$

Existence of a measurable "separator" $g$ such that for general random sources $\mathbf{x}$, $\mathbf{e}$, for sufficiently large blocklengths

$$\mathbb{P}\left[ g\left( [\mathbf{A} \;\; \mathbf{B}] \begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix} \right) \neq \begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix} \right] < \varepsilon$$

$\longrightarrow$ "Almost lossless signal separation"

We are interested in the structure of *pairs* $\mathbf{A}$, $\mathbf{B}$ for which separation is possible. Concretely: fix $\mathbf{B}$, look for suitable $\mathbf{A}$

# Setting

**Source**: $$\big[\underbrace{\mathsf{X}_1\ \cdots\ \mathsf{X}_{n-\ell}}_{\text{fraction: } 1-\lambda}\ \underbrace{\mathsf{E}_1\ \cdots\ \mathsf{E}_\ell}_{\text{fraction: } \lambda}\big]^T \in \mathbb{R}^n$$

stoch. processes: $(\mathsf{X}_i)_{i\in\mathbb{N}}$ and $(\mathsf{E}_i)_{i\in\mathbb{N}}$, fraction parameter: $\lambda \in [0,1]$

# Setting

**Source**: $$\big[\underbrace{\mathsf{X}_1\ \cdots\ \mathsf{X}_{n-\ell}}_{\text{fraction: } 1-\lambda}\ \underbrace{\mathsf{E}_1\ \cdots\ \mathsf{E}_\ell}_{\text{fraction: } \lambda}\big]^T \in \mathbb{R}^n$$

stoch. processes: $(\mathsf{X}_i)_{i\in\mathbb{N}}$ and $(\mathsf{E}_i)_{i\in\mathbb{N}}$, fraction parameter: $\lambda \in [0,1]$

**Code** of rate $R = m/n$:
($m$ = no. of measurements, $n$ = no. of unknowns)

- measurement matrices: $\mathbf{A} \in \mathbb{R}^{m\times(n-\ell)}$, $\mathbf{B} \in \mathbb{R}^{m\times\ell}$
- measurable separator $g\colon \mathbb{R}^m \to \mathbb{R}^{m\times(n-\ell)} \times \mathbb{R}^{m\times\ell}$

# Setting

**Source**: $$\big[\underbrace{\mathsf{X}_1 \;\cdots\; \mathsf{X}_{n-\ell}}_{\text{fraction: } 1-\lambda}\; \underbrace{\mathsf{E}_1 \;\cdots\; \mathsf{E}_\ell}_{\text{fraction: } \lambda}\big]^T \in \mathbb{R}^n$$

stoch. processes: $(\mathsf{X}_i)_{i\in\mathbb{N}}$ and $(\mathsf{E}_i)_{i\in\mathbb{N}}$, fraction parameter: $\lambda \in [0,1]$

**Code** of rate $R = m/n$:
($m =$ no. of measurements, $n =$ no. of unknowns)

- measurement matrices: $\mathbf{A} \in \mathbb{R}^{m\times(n-\ell)}$, $\mathbf{B} \in \mathbb{R}^{m\times\ell}$
- measurable separator $g\colon \mathbb{R}^m \to \mathbb{R}^{m\times(n-\ell)} \times \mathbb{R}^{m\times\ell}$

$R$ is $\varepsilon$-**achievable** if for sufficiently large $n$ (asymptotic analysis)

$$\mathbb{P}\left[g\left([\mathbf{A}\;\;\mathbf{B}]\begin{bmatrix}\mathsf{x}\\ \mathsf{e}\end{bmatrix}\right) \neq \begin{bmatrix}\mathsf{x}\\ \mathsf{e}\end{bmatrix}\right] < \varepsilon$$

# Minkowski Dimension

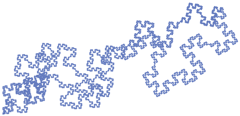
A suitable measure for complexity:

**Covering number**:

$$N_{\mathcal{S}}(\varepsilon) := \min \left\{ k \in \mathbb{N} \;\middle|\; \mathcal{S} \subseteq \bigcup_{i \in \{1,\ldots,k\}} B^n(\boldsymbol{u}_i, \varepsilon),\ \boldsymbol{u}_i \in \mathbb{R}^n \right\}$$

# Minkowski Dimension

A suitable measure for complexity:

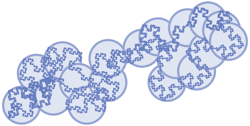


**Covering number**:

$$N_{\mathcal{S}}(\varepsilon) := \min \left\{ k \in \mathbb{N} \;\middle|\; \mathcal{S} \subseteq \bigcup_{i \in \{1,\ldots,k\}} B^n(\boldsymbol{u}_i, \varepsilon),\ \boldsymbol{u}_i \in \mathbb{R}^n \right\}$$

# Minkowski Dimension

A suitable measure for complexity:

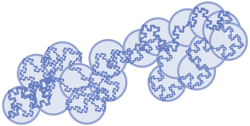


**Covering number**:

$$N_{\mathcal{S}}(\varepsilon) := \min\left\{k \in \mathbb{N} \;\middle|\; \mathcal{S} \subseteq \bigcup_{i\in\{1,\ldots,k\}} B^n(\boldsymbol{u}_i, \varepsilon),\; \boldsymbol{u}_i \in \mathbb{R}^n\right\}$$

# Minkowski Dimension

A suitable measure for complexity:



**Covering number**:

$$N_{\mathcal{S}}(\varepsilon) := \min\left\{k \in \mathbb{N} \;\middle|\; \mathcal{S} \subseteq \bigcup_{i\in\{1,\dots,k\}} B^n(\boldsymbol{u}_i,\varepsilon),\ \boldsymbol{u}_i \in \mathbb{R}^n\right\}$$

**(Lower) Minkowski dimension/Box-counting dimension**:

$$\underline{\dim}_{\mathrm{B}}(\mathcal{S}) := \liminf_{\varepsilon\to 0} \frac{\log N_{\mathcal{S}}(\varepsilon)}{\log \frac{1}{\varepsilon}}$$

$\longrightarrow$ for small $\varepsilon$: $\quad N_{\mathcal{S}}(\varepsilon) \approx \varepsilon^{-\underline{\dim}_{\mathrm{B}}(\mathcal{S})}$

**Minkowski dimension compression rate**:

$$R_{\mathrm{B}}(\varepsilon) := \limsup_{n\to\infty} a_n(\varepsilon) \quad \text{where}$$

$$a_n(\varepsilon) := \inf \left\{ \frac{\overline{\dim}_{\mathrm{B}}(\mathcal{S})}{n} \;\middle|\; \mathcal{S} \subseteq \mathbb{R}^n,\ \mathbb{P}\left[ \begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix} \in \mathcal{S} \right] \geqslant 1 - \varepsilon \right\}$$

# Minkowski Dimension Compression Rate

**Minkowski dimension compression rate**:

$$R_{\mathrm{B}}(\varepsilon) := \limsup_{n\to\infty} a_n(\varepsilon) \quad \text{where}$$

$$a_n(\varepsilon) := \inf\left\{ \frac{\underline{\dim}_{\mathrm{B}}(\mathcal{S})}{n} \;\middle|\; \mathcal{S} \subseteq \mathbb{R}^n,\ \mathbb{P}\left[\begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix} \in \mathcal{S}\right] \geqslant 1-\varepsilon \right\}$$

Among all approximate support sets:

# Minkowski Dimension Compression Rate

**Minkowski dimension compression rate**:

$$R_{\mathrm{B}}(\varepsilon) := \limsup_{n\to\infty} a_n(\varepsilon) \quad \text{where}$$

$$a_n(\varepsilon) := \inf\left\{ \frac{\overline{\dim}_{\mathrm{B}}(\mathcal{S})}{n} \;\middle|\; \mathcal{S} \subseteq \mathbb{R}^n,\ \mathbb{P}\left[\begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix} \in \mathcal{S}\right] \geqslant 1-\varepsilon \right\}$$

Among all approximate support sets:

the smallest possible Minkowski dimension (per blocklength)

# Main Result

**Theorem**

*Let $R > R_{\mathrm{B}}(\varepsilon)$. Then, for every fixed full-rank matrix $\mathbf{B} \in \mathbb{R}^{m\times\ell}$ with $m \geqslant \ell$ and for Lebesgue a.a. matrices $\mathbf{A} \in \mathbb{R}^{m\times(n-\ell)}$, where $m = \lfloor Rn \rfloor$, there exists a measurable separator $g$ such that for sufficiently large $n$*

$$\mathbb{P}\left[ g\left( [\mathbf{A} \;\; \mathbf{B}] \begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix} \right) \neq \begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix} \right] < \varepsilon$$

# Main Result

**Theorem**

*Let $R > R_{\mathrm{B}}(\varepsilon)$. Then, for every fixed full-rank matrix $\mathbf{B} \in \mathbb{R}^{m\times\ell}$ with $m \geqslant \ell$ and for Lebesgue a.a. matrices $\mathbf{A} \in \mathbb{R}^{m\times(n-\ell)}$, where $m = \lfloor Rn \rfloor$, there exists a measurable separator $g$ such that for sufficiently large $n$*

$$\mathbb{P}\left[ g\left( [\mathbf{A} \;\; \mathbf{B}] \begin{bmatrix} \mathsf{x} \\ \mathsf{e} \end{bmatrix} \right) \neq \begin{bmatrix} \mathsf{x} \\ \mathsf{e} \end{bmatrix} \right] < \varepsilon$$

- simple and intuitive proof inspired by [*Sauer et al., 1991*]
- almost all matrices $\mathbf{A}$ are "incoherent" to a given matrix $\mathbf{B}$

# A Probabilistic Null-Space Property

## Proposition

*Let* $\mathcal{S} \subseteq \mathbb{R}^n$ *be such that* $\underline{\dim}_{\mathrm{B}}(\mathcal{S}) < m$ *and let* $\mathbf{B} \in \mathbb{R}^{m \times \ell}$ *be a full-rank matrix with* $m \geqslant \ell$. *Then*

$$\left\{ \boldsymbol{u} \in \mathcal{S} \backslash \{\mathbf{0}\} \;\middle|\; [\mathbf{A} \;\; \mathbf{B}]\boldsymbol{u} = \mathbf{0} \right\} = \emptyset$$

*for Lebesgue a.a.* $\mathbf{A} \in \mathbb{R}^{m \times (n-\ell)}$

# A Probabilistic Null-Space Property

**Proposition**

*Let $\mathcal{S} \subseteq \mathbb{R}^n$ be such that $\underline{\dim}_{\mathrm{B}}(\mathcal{S}) < m$ and let $\mathbf{B} \in \mathbb{R}^{m\times\ell}$ be a full-rank matrix with $m \geqslant \ell$. Then*

$$\left\{\boldsymbol{u} \in \mathcal{S}\backslash\{\mathbf{0}\} \;\middle|\; [\mathbf{A}\ \ \mathbf{B}]\boldsymbol{u} = \mathbf{0}\right\} = \emptyset$$

*for Lebesgue a.a. $\mathbf{A} \in \mathbb{R}^{m\times(n-\ell)}$*

- $\mathbf{A}$ and $\mathbf{B}$ ONBs, then there is no $(\mathbf{p}, \mathbf{q}) \neq 0$ such that

$$\mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q} \quad \text{and} \quad \|\mathbf{p}\|_0 \, \|\mathbf{q}\|_0 < \frac{1}{\mu^2}$$

- $\underline{\dim}_{\mathrm{B}}(\mathcal{S}) < m$, then for a.a. $\mathbf{A}$ there is no $(\mathbf{p}, \mathbf{q}) \neq 0$ such that

$$\mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q} \quad \text{and} \quad \boldsymbol{u} = (\mathbf{p}, -\mathbf{q}) \in \mathcal{S}$$

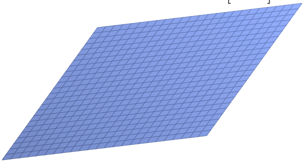
# A Probabilistic Null-Space Property

## Proposition

*Let* $\mathcal{S} \subseteq \mathbb{R}^n$ *be such that* $\underline{\dim}_{\mathrm{B}}(\mathcal{S}) < m$ *and let* $\mathbf{B} \in \mathbb{R}^{m\times\ell}$ *be a full-rank matrix with* $m \geqslant \ell$*. Then*

$$\left\{\boldsymbol{u} \in \mathcal{S}\backslash\{\mathbf{0}\} \;\middle|\; [\mathbf{A}\;\;\mathbf{B}]\boldsymbol{u} = \mathbf{0}\right\} = \emptyset$$

*for Lebesgue a.a.* $\mathbf{A} \in \mathbb{R}^{m\times(n-\ell)}$



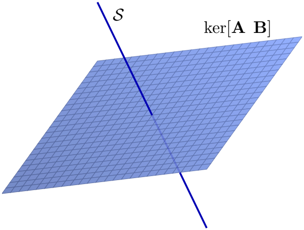
# A Probabilistic Null-Space Property

**Proposition**

*Let $\mathcal{S} \subseteq \mathbb{R}^n$ be such that $\underline{\dim}_{\mathrm{B}}(\mathcal{S}) < m$ and let $\mathbf{B} \in \mathbb{R}^{m\times\ell}$ be a full-rank matrix with $m \geqslant \ell$. Then*

$$\left\{\boldsymbol{u} \in \mathcal{S}\backslash\{\mathbf{0}\} \;\middle|\; [\mathbf{A}\ \ \mathbf{B}]\boldsymbol{u} = \mathbf{0}\right\} = \emptyset$$

*for Lebesgue a.a. $\mathbf{A} \in \mathbb{R}^{m\times(n-\ell)}$*



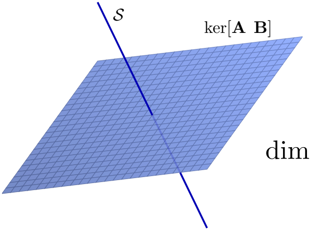
# A Probabilistic Null-Space Property

## Proposition

*Let* $\mathcal{S} \subseteq \mathbb{R}^n$ *be such that* $\underline{\dim}_{\mathrm{B}}(\mathcal{S}) < m$ *and let* $\mathbf{B} \in \mathbb{R}^{m\times\ell}$ *be a full-rank matrix with* $m \geqslant \ell$. *Then*

$$\left\{\boldsymbol{u} \in \mathcal{S}\backslash\{\mathbf{0}\} \;\middle|\; [\mathbf{A}\;\;\mathbf{B}]\boldsymbol{u} = \mathbf{0}\right\} = \emptyset$$

*for Lebesgue a.a.* $\mathbf{A} \in \mathbb{R}^{m\times(n-\ell)}$




$$\dim\ker[\mathbf{A}\;\;\mathbf{B}] = n - m$$

$$\underline{\dim}_{\mathrm{B}}(\mathcal{S}) < m$$

$$\dim\ker[\mathbf{A}\;\;\mathbf{B}] + \underline{\dim}_{\mathrm{B}}(\mathcal{S}) < n$$

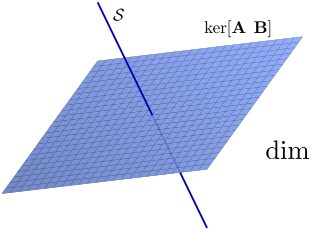
# A Probabilistic Null-Space Property

## Proposition

*Let* $\mathcal{S} \subseteq \mathbb{R}^n$ *be such that* $\underline{\dim}_{\mathrm{B}}(\mathcal{S}) < m$ *and let* $\mathbf{B} \in \mathbb{R}^{m \times \ell}$ *be a full-rank matrix with* $m \geqslant \ell$. *Then*

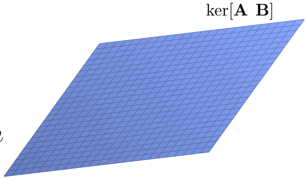
$$\left\{ \boldsymbol{u} \in \mathcal{S} \backslash \{\mathbf{0}\} \;\middle|\; [\mathbf{A} \;\; \mathbf{B}]\boldsymbol{u} = \mathbf{0} \right\} = \emptyset$$

*for Lebesgue a.a.* $\mathbf{A} \in \mathbb{R}^{m \times (n-\ell)}$




$$\dim \ker[\mathbf{A} \;\; \mathbf{B}] = n - m$$

$$\underline{\dim}_{\mathrm{B}}(\mathcal{S}) < m$$




$$\dim \ker[\mathbf{A} \;\; \mathbf{B}] + \underline{\dim}_{\mathrm{B}}(\mathcal{S}) < n$$

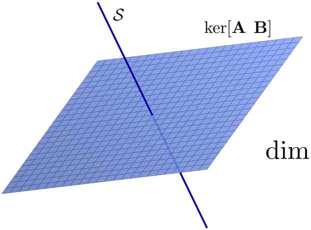
# A Probabilistic Null-Space Property

## Proposition

*Let $\mathcal{S} \subseteq \mathbb{R}^n$ be such that $\underline{\dim}_{\mathrm{B}}(\mathcal{S}) < m$ and let $\mathbf{B} \in \mathbb{R}^{m\times\ell}$ be a full-rank matrix with $m \geqslant \ell$. Then*

$$\left\{\boldsymbol{u} \in \mathcal{S}\backslash\{\mathbf{0}\} \;\middle|\; [\mathbf{A}\ \ \mathbf{B}]\boldsymbol{u} = \mathbf{0}\right\} = \emptyset$$

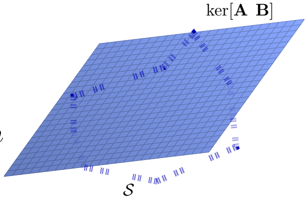
*for Lebesgue a.a. $\mathbf{A} \in \mathbb{R}^{m\times(n-\ell)}$*




$$\dim\ker[\mathbf{A}\ \ \mathbf{B}] = n - m$$

$$\underline{\dim}_{\mathrm{B}}(\mathcal{S}) < m$$

$$\dim\ker[\mathbf{A}\ \ \mathbf{B}] + \underline{\dim}_{\mathrm{B}}(\mathcal{S}) < n$$



# Back to Discrete-Continuous Mixtures

$$\begin{aligned} \mathsf{X}_i \ \text{ i.i.d.} &\sim (1-\rho_1)P_{\mathsf{d}_1} + \rho_1 P_{\mathsf{c}_1} \\ \mathsf{E}_i \ \text{ i.i.d.} &\sim (1-\rho_2)P_{\mathsf{d}_2} + \rho_2 P_{\mathsf{c}_2} \end{aligned}$$

where $0 \leqslant \rho_i \leqslant 1$; for $P_{\mathsf{d}_1} = P_{\mathsf{d}_2} = \delta_0 \rightarrow$ sparse signal model
Fraction of $\mathsf{X}_i$'s $= 1 - \lambda$; fraction of $\mathsf{E}_i$'s $= \lambda$

# Back to Discrete-Continuous Mixtures

$$\mathsf{X}_i \ \text{i.i.d.} \ \sim (1-\rho_1)P_{\mathsf{d}_1} + \rho_1 P_{\mathsf{c}_1}$$

$$\mathsf{E}_i \ \text{i.i.d.} \ \sim (1-\rho_2)P_{\mathsf{d}_2} + \rho_2 P_{\mathsf{c}_2}$$

where $0 \leqslant \rho_i \leqslant 1$; for $P_{\mathsf{d}_1} = P_{\mathsf{d}_2} = \delta_0 \to$ sparse signal model
Fraction of $\mathsf{X}_i$'s $= 1-\lambda$; fraction of $\mathsf{E}_i$'s $= \lambda$

An **exact expression** for $R_{\mathrm{B}}(\varepsilon)$ and a **converse**:

**Theorem**

*For discrete-continuous mixtures the optimal compression rate is*

$$R_{\mathrm{B}}(\varepsilon) = (1-\lambda)\rho_1 + \lambda\rho_2$$

# Back to Discrete-Continuous Mixtures

$$\begin{aligned} \mathsf{X}_i \ \text{ i.i.d.} \ &\sim (1-\rho_1)P_{\mathsf{d}_1} + \rho_1 P_{\mathsf{c}_1} \\ \mathsf{E}_i \ \text{ i.i.d.} \ &\sim (1-\rho_2)P_{\mathsf{d}_2} + \rho_2 P_{\mathsf{c}_2} \end{aligned}$$

where $0 \leqslant \rho_i \leqslant 1$; for $P_{\mathsf{d}_1} = P_{\mathsf{d}_2} = \delta_0 \rightarrow$ sparse signal model
Fraction of $\mathsf{X}_i$'s $= 1-\lambda$; fraction of $\mathsf{E}_i$'s $= \lambda$

An **exact expression** for $R_{\mathrm{B}}(\varepsilon)$ and a **converse**:

**Theorem**

*For discrete-continuous mixtures the optimal compression rate is*

$$R_{\mathrm{B}}(\varepsilon) = (1-\lambda)\rho_1 + \lambda\rho_2$$

Optimal no. of measurements = no. of nonzero components

# What it is and What it is not

The information-theoretic approach

- applies to general source distributions

# What it is and What it is not

The information-theoretic approach

- applies to general source distributions
- works for given $\mathbf{B}$ for a.a. $\mathbf{A}$

# What it is and What it is not

The information-theoretic approach

- applies to general source distributions
- works for given $\mathbf{B}$ for a.a. $\mathbf{A}$
- yields explicit expression for optimal compression rate (i.e., optimal no. of measurements) for sparse signals

# What it is and What it is not

The information-theoretic approach

- applies to general source distributions
- works for given $\mathbf{B}$ for a.a. $\mathbf{A}$
- yields explicit expression for optimal compression rate (i.e., optimal no. of measurements) for sparse signals

  $\longrightarrow$ achieves linear scaling (got rid of $\log n$ factor)

# What it is and What it is not

The information-theoretic approach

- applies to general source distributions
- works for given $\mathbf{B}$ for a.a. $\mathbf{A}$
- yields explicit expression for optimal compression rate (i.e., optimal no. of measurements) for sparse signals
  $\longrightarrow$ achieves linear scaling (got rid of $\log n$ factor)
- is of asymptotic nature

# What it is and What it is not

The information-theoretic approach

- applies to general source distributions
- works for given $\mathbf{B}$ for a.a. $\mathbf{A}$
- yields explicit expression for optimal compression rate (i.e., optimal no. of measurements) for sparse signals
  $\longrightarrow$ achieves linear scaling (got rid of $\log n$ factor)
- is of asymptotic nature
- deals with the noiseless case

# What it is and What it is not

The information-theoretic approach

- applies to general source distributions
- works for given $\mathbf{B}$ for a.a. $\mathbf{A}$
- yields explicit expression for optimal compression rate (i.e., optimal no. of measurements) for sparse signals
  $\longrightarrow$ achieves linear scaling (got rid of $\log n$ factor)
- is of asymptotic nature
- deals with the noiseless case
- provides existence results only for decoders

Thank you

“If you ask me anything I don’t know,

I’m not going to answer.”

— Y. Berra