Replica Method for Sparse Representation of White Gaussian Noise with Application to Noisy Compressed Sensing

Ori Shental

Qualcomm Inc.

Overview

- WGN is non-compressible.

Overview

- WGN is non-compressible.
- However WGN realizations do have sparse representations

Overview

- WGN is non-compressible.
- However WGN realizations do have sparse representations
- Unresolved questions:

Overview

- WGN is non-compressible.
- However WGN realizations do have sparse representations
- Unresolved questions:
  - What is the minimal Hamming weight of WGN sparse representations?

Overview

- WGN is non-compressible.
- However WGN realizations do have sparse representations
- Unresolved questions:
  - What is the minimal Hamming weight of WGN sparse representations?
  - What are the achievable and converse regions of WGN sparse representation?

Overview

- WGN is non-compressible.
- However WGN realizations do have sparse representations
- Unresolved questions:
  - What is the minimal Hamming weight of WGN sparse representations?
  - What are the achievable and converse regions of WGN sparse representation?
  - What is the probabilistic profile of such representations?

Overview

- WGN is non-compressible.
- However WGN realizations do have sparse representations
- Unresolved questions:
  - What is the minimal Hamming weight of WGN sparse representations?
  - What are the achievable and converse regions of WGN sparse representation?
  - What is the probabilistic profile of such representations?
- These fundamental questions are answered in this work via the Replica method, leading to:

Overview

- WGN is non-compressible.
- However WGN realizations do have sparse representations
- Unresolved questions:
  - What is the minimal Hamming weight of WGN sparse representations?
  - What are the achievable and converse regions of WGN sparse representation?
  - What is the probabilistic profile of such representations?
- These fundamental questions are answered in this work via the Replica method, leading to:
  - Sharp threshold for ℓ_0-norm decoding in noisy compressed sensing (viz. underdetermined Gaussian vector channels)

Overview

- WGN is non-compressible.
- However WGN realizations do have sparse representations
- Unresolved questions:
  - What is the minimal Hamming weight of WGN sparse representations?
  - What are the achievable and converse regions of WGN sparse representation?
  - What is the probabilistic profile of such representations?
- These fundamental questions are answered in this work via the Replica method, leading to:
  - Sharp threshold for ℓ_0-norm decoding in noisy compressed sensing (viz. underdetermined Gaussian vector channels)
  - and its mean-square error performance

Acknowledgment for useful discussions

- Ido Kanter (BIU)
- Noam Shental (Open U.)
- Dror Baron (NCSU)
- Toshiyuki Tanaka (Kyoto U.)
- Andrea Montanari (Stanford)
- Rick Chartrand (LANL)

Outline

Introduction

Sparse representation of WGN

Noisy compressed sensing

Outline

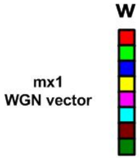
Introduction

Sparse representation of WGN

Noisy compressed sensing

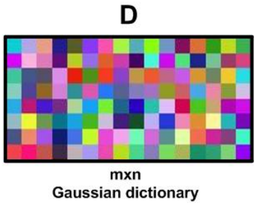
Definitions

- Let $\mathbf{w} \in \mathbb{R}^m$ be a WGN vector with i.i.d. $w_i \sim \mathcal{N}(0, 1)$, $i = 1, \ldots, m$
- Choose an overcomplete dictionary $\mathbf{D} \in \mathbb{R}^{m \times n}$, $m \leq n$, with zero-mean, unit-variance, i.i.d. D_{ij}, $j = 1, \ldots, n$
  - *e.g.* Gaussian $D_{ij} \sim \mathcal{N}(0, 1)$, Bernoulli $D_{ij} = \pm 1$
- $\mathbf{w}$ and $\mathbf{D}$ are statistically independent
- The realizations of the WGN, $\mathbf{w}$, and dictionary, $\mathbf{D}$, are denoted by ω and $\mathcal{D}$, respectively.



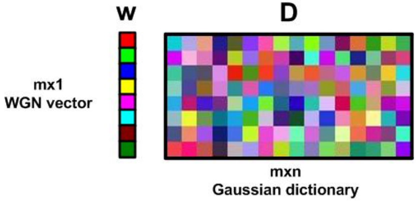
Definitions

- Let $\mathbf{w} \in \mathbb{R}^m$ be a WGN vector with i.i.d. $w_i \sim \mathcal{N}(0,1)$, $i = 1, \ldots, m$
- Choose an overcomplete dictionary $\mathbf{D} \in \mathbb{R}^{m \times n}$, $m \leq n$, with zero-mean, unit-variance, i.i.d. D_{ij}, $j = 1, \ldots, n$
  - *e.g.* Gaussian $D_{ij} \sim \mathcal{N}(0,1)$, Bernoulli $D_{ij} = \pm 1$
- $\mathbf{w}$ and $\mathbf{D}$ are statistically independent
- The realizations of the WGN, $\mathbf{w}$, and dictionary, $\mathbf{D}$, are denoted by ω and $\mathcal{D}$, respectively.



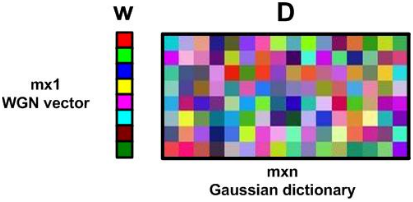
Definitions

- Let $\mathbf{w} \in \mathbb{R}^m$ be a WGN vector with i.i.d. $w_i \sim \mathcal{N}(0, 1)$, $i = 1, \ldots, m$
- Choose an overcomplete dictionary $\mathbf{D} \in \mathbb{R}^{m \times n}$, $m \leq n$, with zero-mean, unit-variance, i.i.d. D_{ij}, $j = 1, \ldots, n$
  - *e.g.* Gaussian $D_{ij} \sim \mathcal{N}(0, 1)$, Bernoulli $D_{ij} = \pm 1$
- $\mathbf{w}$ and $\mathbf{D}$ are statistically independent
- The realizations of the WGN, $\mathbf{w}$, and dictionary, $\mathbf{D}$, are denoted by ω and $\mathcal{D}$, respectively.



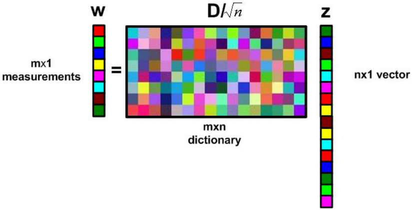
Definitions

- Let $\mathbf{w} \in \mathbb{R}^m$ be a WGN vector with i.i.d. $w_i \sim \mathcal{N}(0,1)$, $i = 1, \ldots, m$
- Choose an overcomplete dictionary $\mathbf{D} \in \mathbb{R}^{m \times n}$, $m \leq n$, with zero-mean, unit-variance, i.i.d. D_{ij}, $j = 1, \ldots, n$
  - *e.g.* Gaussian $D_{ij} \sim \mathcal{N}(0,1)$, Bernoulli $D_{ij} = \pm 1$
- $\mathbf{w}$ and $\mathbf{D}$ are statistically independent
- The realizations of the WGN, $\mathbf{w}$, and dictionary, $\mathbf{D}$, are denoted by $\boldsymbol{\omega}$ and $\mathcal{D}$, respectively.



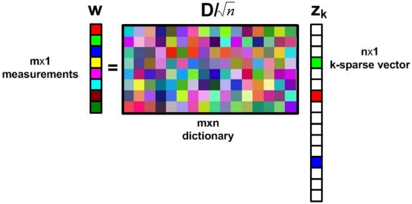
Definitions

- Let $\mathsf{z}(\omega, \mathcal{D}) \in \mathbb{R}^n$ be a representation of a WGN vector via a dictionary, namely
$$\omega = \frac{1}{\sqrt{n}}\mathcal{D}\mathsf{z}$$
- $\mathsf{z}_k(\omega, \mathcal{D})$ is termed k-sparse representation if at most $k \leq n$ of its entries are non-zero



Definitions

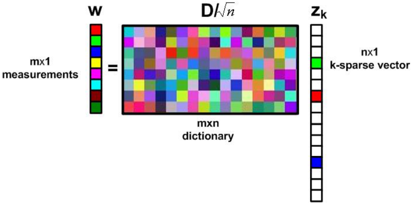
- Let $\mathbf{z}(\boldsymbol{\omega}, \boldsymbol{\mathcal{D}}) \in \mathbb{R}^n$ be a representation of a WGN vector via a dictionary, namely
$$\boldsymbol{\omega} = \frac{1}{\sqrt{n}} \boldsymbol{\mathcal{D}} \mathbf{z}$$
- $\mathbf{z}_k(\boldsymbol{\omega}, \boldsymbol{\mathcal{D}})$ is termed k-sparse representation if at most $k \leq n$ of its entries are non-zero



Large-system limit

Our analysis assumes:

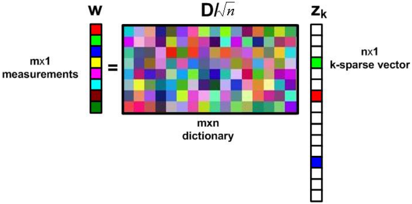
- $m, n, k \to \infty$
- Measurement ratio $\alpha \triangleq m/n \in (0, 1)$
- Sparsity fraction $\kappa \triangleq k/n \in (0, 1)$



Large-system limit

Our analysis assumes:

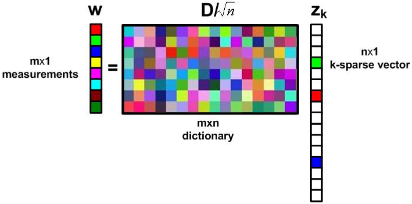
- $m, n, k \to \infty$
- Measurement ratio $\alpha \triangleq m/n \in (0,1)$
- Sparsity fraction $\kappa \triangleq k/n \in (0,1)$



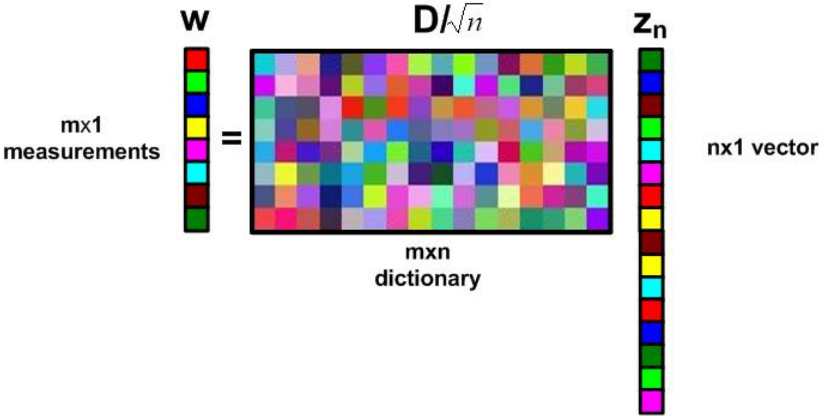
Large-system limit

Our analysis assumes:

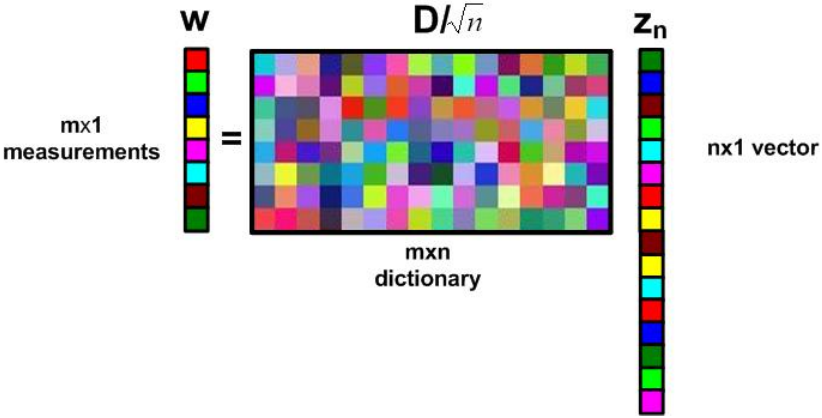
- $m, n, k \to \infty$
- Measurement ratio $\alpha \triangleq m/n \in (0, 1)$
- Sparsity fraction $\kappa \triangleq k/n \in (0, 1)$



Example: $(\kappa = 1)$-sparse (dense) representation



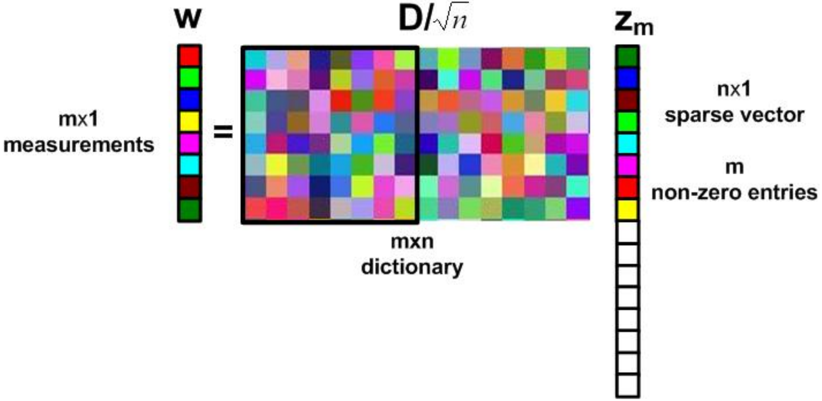
Example: $(\kappa = 1)$-sparse (dense) representation



- For example:
$\mathbf{z}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \Rightarrow \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}\mathbf{D}^T/n) \xrightarrow[\alpha\in(0,1)]{m,n\to\infty} \mathcal{N}(\mathbf{0}, \mathbf{I})$

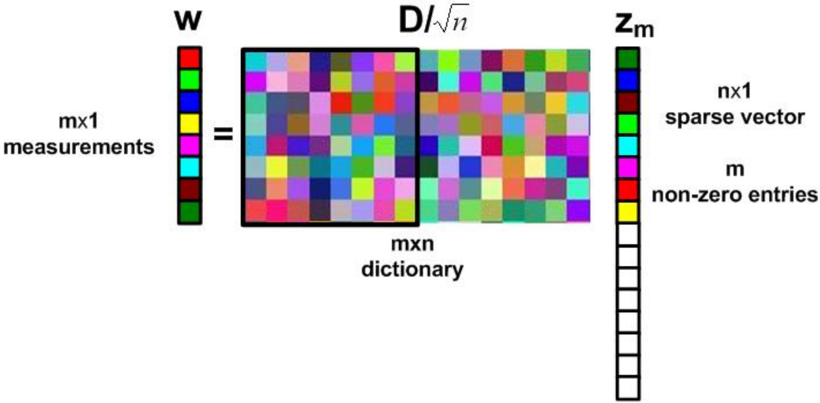
Another example: $(\kappa = \alpha)$-sparse representation

- Trivially, one can also represent (almost surely) the WGN vector using only m out of n entries



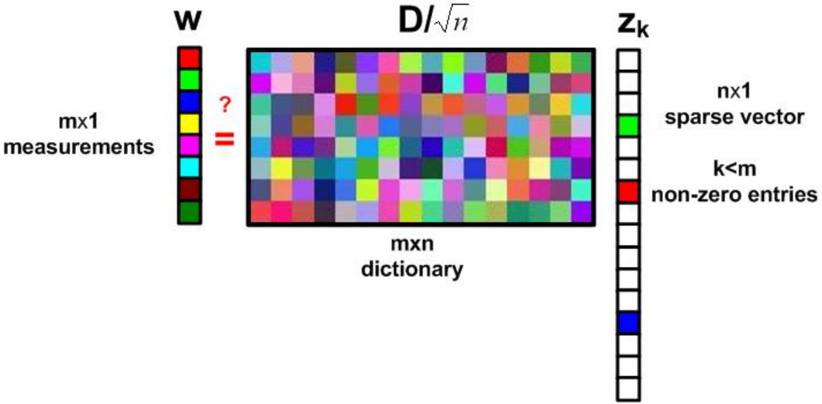
Motivation

- But can we go below this trivial sparsity?



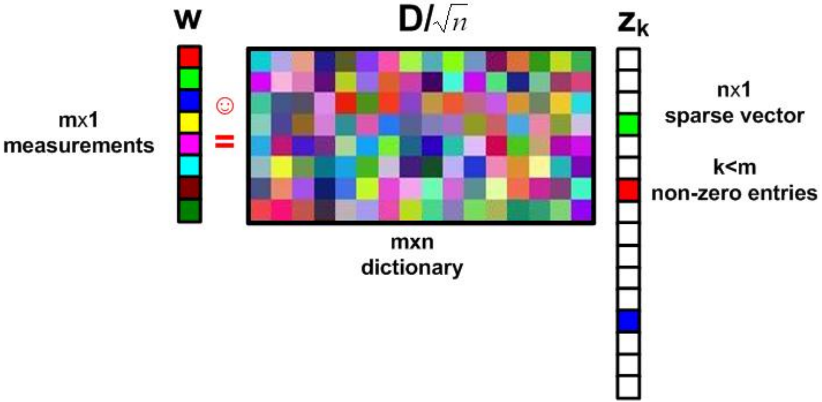
Motivation

- But can we go below this trivial sparsity?



Motivation

- Intriguingly the answer is yes: as we shall see, there are sparser representations of WGN with some interesting characteristics



Problem formulation

- What is the normalized Hamming weight (sparsity fraction) $\kappa_\alpha^*(\boldsymbol{\omega}, \boldsymbol{\mathcal{D}})$ of the sparsest representation of WGN based on an α-measurement dictionary in the limit of large systems?

Problem formulation

- What is the normalized Hamming weight (sparsity fraction) $\kappa_\alpha^*(\boldsymbol{\omega}, \boldsymbol{\mathcal{D}})$ of the $\underline{\text{sparsest}}$ representation of WGN based on an α-measurement dictionary in the limit of large systems?

$$\mathsf{z}_{\kappa_\alpha^*}(\boldsymbol{\omega}, \boldsymbol{\mathcal{D}}) = \arg\min \|\mathsf{z}\|_0 \quad \text{subject to} \quad \boldsymbol{\omega} = \frac{1}{\sqrt{n}} \boldsymbol{\mathcal{D}} \mathsf{z}$$

where $\|\mathsf{z}\|_0 \triangleq \#\{i \in \{1, \ldots, n\} \mid z_i \neq 0\}$

Problem reformulation

- What is the normalized Hamming weight (sparsity fraction) κ_α^* of the sparsest representation of WGN based on an α-measurement dictionary in the limit of large systems?

$$\mathbf{z}_{\kappa_\alpha^*}(\boldsymbol{\omega}, \boldsymbol{\mathcal{D}}) = \arg\min \|\mathbf{z}\|_0 \quad \text{subject to} \quad \boldsymbol{\omega} = \frac{1}{\sqrt{n}} \boldsymbol{\mathcal{D}} \mathbf{z}$$

Problem reformulation

- What is the normalized Hamming weight (sparsity fraction) κ_α^* of the sparsest representation of WGN based on an α-measurement dictionary in the limit of large systems?

$$\mathbf{z}_{\kappa_\alpha^*}(\boldsymbol{\omega}, \boldsymbol{\mathcal{D}}) = \arg\min \|\mathbf{z}\|_0 \quad \text{subject to} \quad \boldsymbol{\omega} = \frac{1}{\sqrt{n}}\boldsymbol{\mathcal{D}}\mathbf{z}$$

Assumption (Self-averaging property)

The limit $\kappa_\alpha^* \triangleq \lim_{n\to\infty} \kappa_\alpha^*(\boldsymbol{\omega}, \boldsymbol{\mathcal{D}})$ *exists and it is equal to its average over the randomness of the WGN and dictionary,* $\lim_{n\to\infty} \mathbb{E}_{\mathbf{w},\mathbf{D}}\{\kappa_\alpha^*(\boldsymbol{\omega}, \boldsymbol{\mathcal{D}})\}$*, for almost all realizations of the WGN and dictionary.*

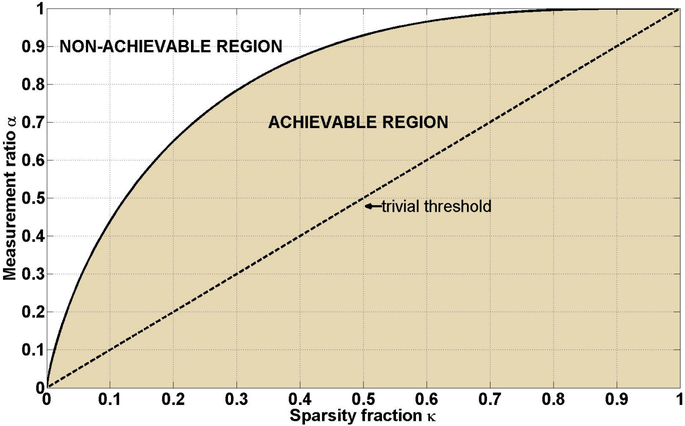
Outline

Introduction

Sparse representation of WGN

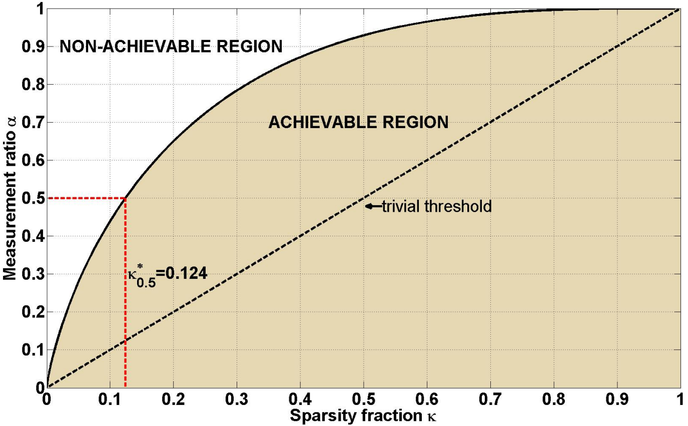
Noisy compressed sensing

Illustration



- Introducing the achievable and converse regions for sparse representation of WGN

Illustration



- Introducing the achievable and converse regions for sparse representation of WGN

Proposition (Sparsest representation of WGN)

Consider the scalars $\alpha \in (0,1)$ *and* $\kappa_\alpha^* \triangleq 2\mathcal{Q}(\xi) \in (0,1)$*, where* $\xi \geq 0$ *is determined by*

$$\alpha = \sqrt{\frac{2}{\pi}} \int_\xi^\infty t^2 \exp(-t^2/2) dt,$$

and $\mathcal{Q}(\xi) \triangleq \int_\xi^\infty dt/\sqrt{2\pi} \exp(-t^2/2)$ *is the Q-function.*
Then, with probability 1 in the large-system limit, for a dictionary **D** *with measurement ratio* α *:*

i) *minimal Hamming weight: the sparsest WGN representation is* κ_α^**-sparse;*

ii) *achievable region:* κ*-sparse representation of WGN exists only for* $\kappa \geq \kappa_\alpha^*$*;*

iii) *converse region:* κ*-sparse representation of WGN does not exist for* $\kappa < \kappa_\alpha^*$*.*

Sparsest representation of WGN

- Proof is based on the Replica method from statistical mechanics (Replica-symmetric ansatz)
- Replica method is mathematically non-rigorous, but shown to be very powerful tool

Achieving distribution of WGN sparse representation

- What is the probability density function of the WGN κ-sparse representation via α-measurement dictionary?

Achieving distribution of WGN sparse representation

Proposition

The marginal probability density function of the j'th (non-zero) entry, ζ_j, of the (minimal ℓ_2-norm) κ-sparse representation of WGN, $\mathbf{z}_\kappa$, is given in the large-system limit by

$$p(\zeta_j) = \begin{cases} 0 & \textit{if } |\zeta_j| < \xi \sqrt{\frac{\alpha}{\alpha_\kappa^*(\alpha_\kappa^* - \alpha)}} \\ \sqrt{\frac{\alpha_\kappa^*(\alpha_\kappa^* - \alpha)}{2\pi\alpha\kappa^2}} \exp\left(- \frac{\zeta_j^2 \alpha_\kappa^*(\alpha_\kappa^* - \alpha)}{2\alpha} \right) & \textit{otherwise} \end{cases},$$

where α_κ^* is the achievability threshold for given sparsity fraction κ.

Achieving distribution of WGN sparse representation

Proposition

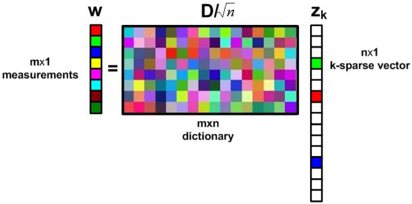
The marginal probability density function of the j'th (non-zero) entry, ζ_j, of the (minimal ℓ_2-norm) κ-sparse representation of WGN, $\mathbf{z}_\kappa$, is given in the large-system limit by

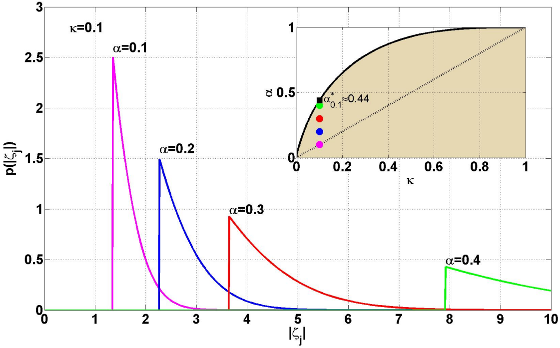
$$p(\zeta_j) = \begin{cases} 0 & \text{if } |\zeta_j| < \xi \sqrt{\frac{\alpha}{\alpha^*_\kappa(\alpha^*_\kappa - \alpha)}} \\ \sqrt{\frac{\alpha^*_\kappa(\alpha^*_\kappa - \alpha)}{2\pi\alpha\kappa^2}} \exp\left(-\frac{\zeta_j^2 \alpha^*_\kappa(\alpha^*_\kappa - \alpha)}{2\alpha}\right) & \text{otherwise} \end{cases},$$

where α^*_κ is the achievability threshold for given sparsity fraction κ.

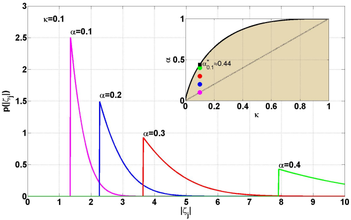
- Proof is also based on Replica analysis

Example: sparsity fraction $\kappa = 0.1$



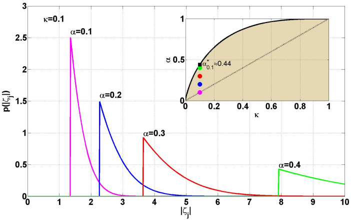


Example: sparsity fraction $\kappa = 0.1$



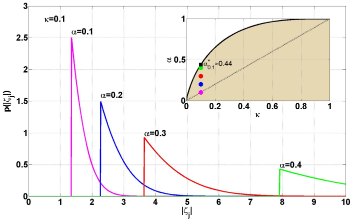
- $\alpha \to \alpha^* \Rightarrow$ The gap increases to infinity and the non-vanishing part (Gaussian tail) of the distribution becomes more uniform
- The marginal probability is symmetric
- In the large-system limit, the achieving distribution is <u>not</u> a function of the WGN and Gaussian dictionary realizations
- There is an infinite number of sparse representations per (κ, α)-point in the achievable region
- The stated achieving distribution corresponds to the minimal ℓ_2-norm representation
- What is the pairwise/joint distribution of ζ? interesting (yet challenging) open question

Example: sparsity fraction $\kappa = 0.1$



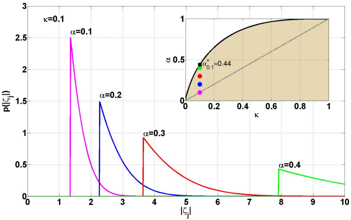
- $\alpha \to \alpha^* \Rightarrow$ The gap increases to infinity and the non-vanishing part (Gaussian tail) of the distribution becomes more uniform
- The marginal probability is symmetric
- In the large-system limit, the achieving distribution is not a function of the WGN and Gaussian dictionary realizations
- There is an infinite number of sparse representations per (κ, α)-point in the achievable region
- The stated achieving distribution corresponds to the minimal ℓ_2-norm representation
- What is the pairwise/joint distribution of ζ? interesting (yet challenging) open question

Example: sparsity fraction $\kappa = 0.1$



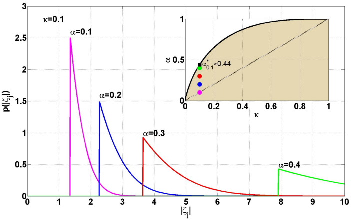
- $\alpha \to \alpha^* \Rightarrow$ The gap increases to infinity and the non-vanishing part (Gaussian tail) of the distribution becomes more uniform
- The marginal probability is symmetric
- In the large-system limit, the achieving distribution is <u>not</u> a function of the WGN and Gaussian dictionary realizations
- There is an infinite number of sparse representations per (κ, α)-point in the achievable region
- The stated achieving distribution corresponds to the minimal ℓ_2-norm representation
- What is the pairwise/joint distribution of ζ? interesting (yet challenging) open question

Example: sparsity fraction $\kappa = 0.1$



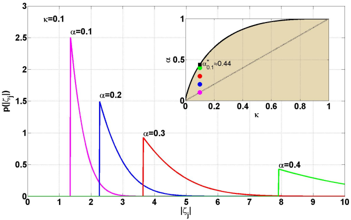
- $\alpha \to \alpha^* \Rightarrow$ The gap increases to infinity and the non-vanishing part (Gaussian tail) of the distribution becomes more uniform
- The marginal probability is symmetric
- In the large-system limit, the achieving distribution is not a function of the WGN and Gaussian dictionary realizations
- There is an infinite number of sparse representations per (κ, α)-point in the achievable region
- The stated achieving distribution corresponds to the minimal ℓ_2-norm representation
- What is the pairwise/joint distribution of ζ? interesting (yet challenging) open question

Example: sparsity fraction $\kappa = 0.1$



- $\alpha \to \alpha^* \Rightarrow$ The gap increases to infinity and the non-vanishing part (Gaussian tail) of the distribution becomes more uniform
- The marginal probability is symmetric
- In the large-system limit, the achieving distribution is not a function of the WGN and Gaussian dictionary realizations
- There is an infinite number of sparse representations per (κ, α)-point in the achievable region
- The stated achieving distribution corresponds to the minimal ℓ_2-norm representation
- What is the pairwise/joint distribution of ζ? interesting (yet challenging) open question

Example: sparsity fraction $\kappa = 0.1$



- $\alpha \to \alpha^* \Rightarrow$ The gap increases to infinity and the non-vanishing part (Gaussian tail) of the distribution becomes more uniform
- The marginal probability is symmetric
- In the large-system limit, the achieving distribution is not a function of the WGN and Gaussian dictionary realizations
- There is an infinite number of sparse representations per (κ, α)-point in the achievable region
- The stated achieving distribution corresponds to the minimal ℓ_2-norm representation
- What is the pairwise/joint distribution of ζ? interesting (yet challenging) open question

Experimental study: setup

- Generate WGN vector `w=randn(m,1)` and Gaussian dictionary `D=randn(m,n)`
- Estimate the sparsest representation per $\mathbf{w}$ and $\mathbf{D}$ realization using the iteratively reweighted least-squares (IRLS) method, which is an approximation of ℓ_0-norm minimization
- Average over realizations

Experimental study: setup

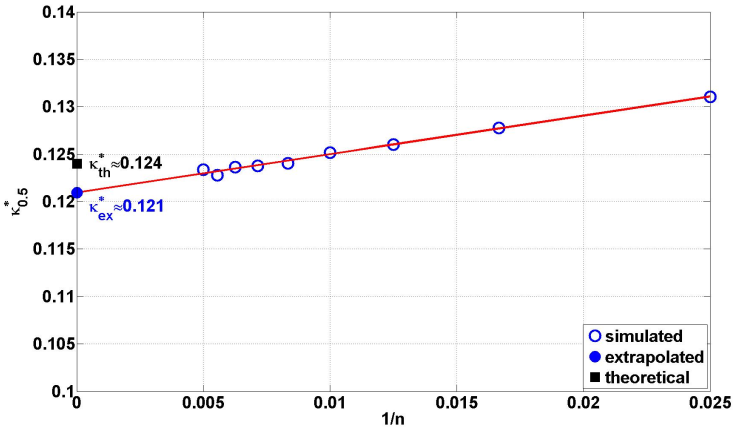
- Generate WGN vector `w=randn(m,1)` and Gaussian dictionary `D=randn(m,n)`
- Estimate the sparsest representation per $\mathbf{w}$ and $\mathbf{D}$ realization using the iteratively reweighted least-squares (IRLS) method, which is an approximation of ℓ_0-norm minimization
- Average over realizations

Experimental study: setup

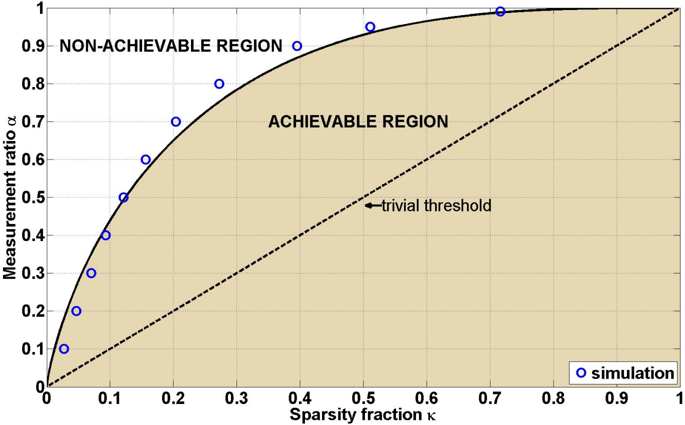
- Generate WGN vector `w=randn(m,1)` and Gaussian dictionary `D=randn(m,n)`
- Estimate the sparsest representation per $\mathbf{w}$ and $\mathbf{D}$ realization using the iteratively reweighted least-squares (IRLS) method, which is an approximation of ℓ_0-norm minimization
- Average over realizations

Experimental study: results

- Quadratic curve fitting is used to extrapolate threshold for $n \to \infty$
- For example:



Experimental study: results



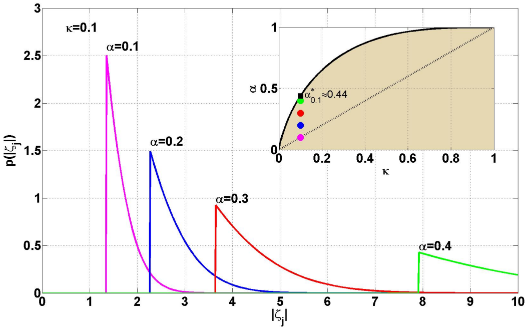
- Fairly good agreement
- Mismatch due to extrapolation errors and IRLS approximation

Experimental study: setup

- Generate Gaussian dictionary `D=randn(m,n)`
- Generate sparse representation $\mathbf{z}$ according to the Proposition (non-zero entry locations are chosen uniformly)
- Generate $\mathbf{w} = \frac{1}{\sqrt{n}}\mathbf{Dz}$
- Build histogram for $\mathbf{w}$

Experimental study: setup

- Generate Gaussian dictionary `D=randn(m,n)`
- Generate sparse representation $\mathbf{z}$ according to the Proposition (non-zero entry locations are chosen uniformly)
- Generate $\mathbf{w} = \frac{1}{\sqrt{n}}\mathbf{D}\mathbf{z}$
- Build histogram for $\mathbf{w}$



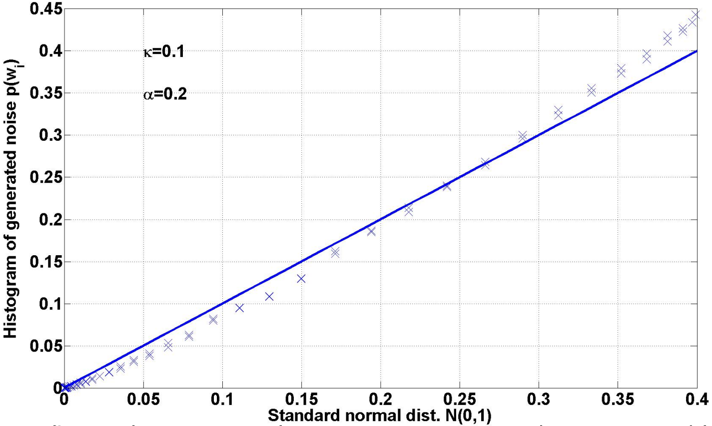
Experimental study: setup

- Generate Gaussian dictionary `D=randn(m,n)`
- Generate sparse representation $\mathbf{z}$ according to the Proposition (non-zero entry locations are chosen uniformly)
- Generate $\mathbf{w} = \frac{1}{\sqrt{n}}\mathbf{D}\mathbf{z}$
- Build histogram for $\mathbf{w}$

Experimental study: setup

- Generate Gaussian dictionary `D=randn(m,n)`
- Generate sparse representation $\mathbf{z}$ according to the Proposition (non-zero entry locations are chosen uniformly)
- Generate $\mathbf{w} = \frac{1}{\sqrt{n}}\mathbf{D}\mathbf{z}$
- Build histogram for $\mathbf{w}$

Experimental study: results



- Reflects the generated noise is approximately Gaussian (does not say anything about it being white)
- Mismatch is mainly due to the inaccurate assumption that the non-zero entries of the sparse representation, z_i, are i.i.d.

So what we have learned so far?

- The minimal Hamming weight of sparse representation of WGN
- The marginal distribution of such sparse representations

Outline

Introduction

Sparse representation of WGN

Noisy compressed sensing

Noiseless compressed sensing

- Let $\mathbf{x}_k$ be a k-sparse n-length vector
- Underdetermined linear transformation $\mathcal{D}\mathbf{x}_k$ is observed
- Reconstruction with ℓ_0-norm

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathcal{D}\mathbf{x}_k = \mathcal{D}\mathbf{x}$$

Noiseless compressed sensing

- Let $\mathbf{x}_k$ be a k-sparse n-length vector
- Underdetermined linear transformation $\mathcal{D}\mathbf{x}_k$ is observed
- Reconstruction with ℓ_0-norm

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathcal{D}\mathbf{x}_k = \mathcal{D}\mathbf{x}$$

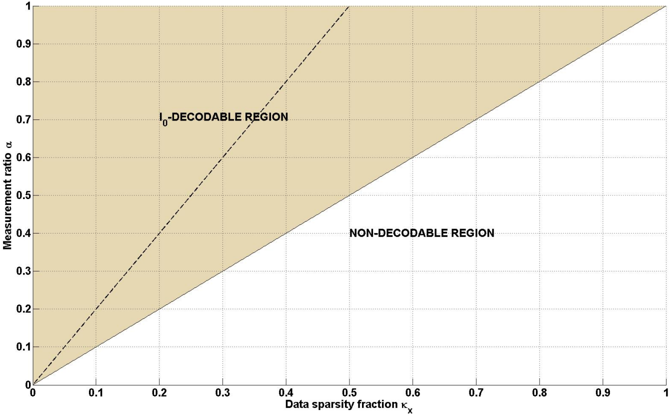
Noiseless compressed sensing

- Let $\mathbf{x}_k$ be a k-sparse n-length vector
- Underdetermined linear transformation $\mathcal{D}\mathbf{x}_k$ is observed
- Reconstruction with ℓ_0-norm

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathcal{D}\mathbf{x}_k = \mathcal{D}\mathbf{x}$$

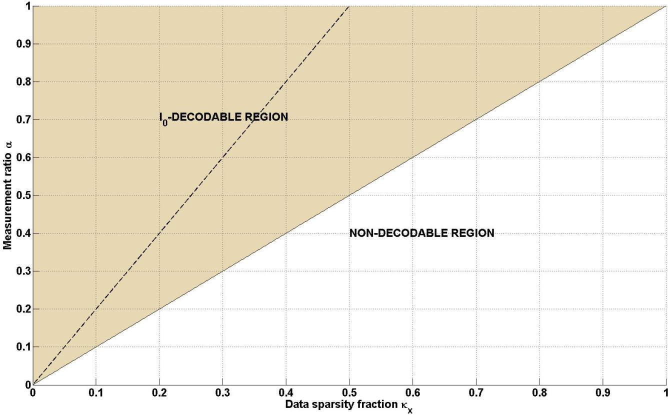
Noiseless compressed sensing

- Prefect reconstruction $\hat{\mathbf{x}} = \mathbf{x}_k$ is possible with probability 1 if
  - $m \geq k + 1$ for almost any $\mathbf{x}_k$ (weak achievable region)
  - $m \geq 2k$ for any $\mathbf{x}_k$ (strong achievable region)
- Reconstruction is impossible with overwhelming probability for $m \leq k$ (strong converse region)



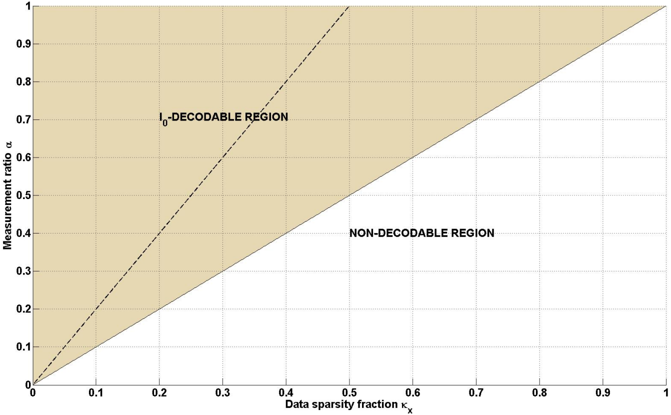
Noiseless compressed sensing

- Prefect reconstruction $\hat{\mathbf{x}} = \mathbf{x}_k$ is possible with probability 1 if
  - $m \geq k + 1$ for almost any $\mathbf{x}_k$ (weak achievable region)
  - $m \geq 2k$ for any $\mathbf{x}_k$ (strong achievable region)
- Reconstruction is impossible with overwhelming probability for $m \leq k$ (strong converse region)



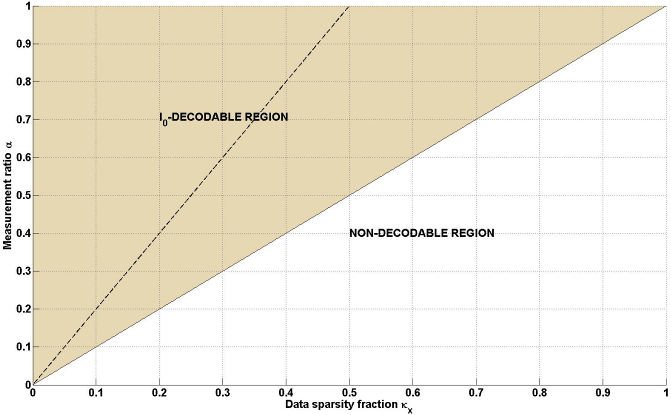
Noiseless compressed sensing

- Prefect reconstruction $\hat{\mathbf{x}} = \mathbf{x}_k$ is possible with probability 1 if
  - $m \geq k + 1$ for almost any $\mathbf{x}_k$ (weak achievable region)
  - $m \geq 2k$ for any $\mathbf{x}_k$ (strong achievable region)
- Reconstruction is impossible with overwhelming probability for $m \leq k$ (strong converse region)



Noiseless compressed sensing

- Prefect reconstruction $\hat{\mathbf{x}} = \mathbf{x}_k$ is possible with probability 1 if
  - $m \geq k + 1$ for almost any $\mathbf{x}_k$ (weak achievable region)
  - $m \geq 2k$ for any $\mathbf{x}_k$ (strong achievable region)
- Reconstruction is impossible with overwhelming probability for $m \leq k$ (strong converse region)



ℓ_0-norm optimization complexity

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathcal{D}\mathbf{x}_k = \mathcal{D}\mathbf{x}$$

- Prohibitively complex optimization problem
  - *NP*-complete: requires combinatorial enumeration of the $\binom{n}{k}$ possible sparse vectors
  - Still of theoretical importance and may lead to practical consequences and insights (*e.g.*, bounds performance of tractable reconstruction methods, like $\ell_1 - norm$ optimization)

ℓ_0-norm optimization complexity

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_0 \quad \text{subject to} \quad \boldsymbol{\mathcal{D}}\mathbf{x}_k = \boldsymbol{\mathcal{D}}\mathbf{x}$$

- Prohibitively complex optimization problem
  - *NP*-complete: requires combinatorial enumeration of the $\binom{n}{k}$ possible sparse vectors
  - Still of theoretical importance and may lead to practical consequences and insights (*e.g.*, bounds performance of tractable reconstruction methods, like $\ell_1 - \textit{norm}$ optimization)

ℓ_0-norm optimization complexity

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_0 \quad \text{subject to} \quad \boldsymbol{\mathcal{D}}\mathbf{x}_k = \boldsymbol{\mathcal{D}}\mathbf{x}$$

- Prohibitively complex optimization problem
  - *NP*-complete: requires combinatorial enumeration of the $\binom{n}{k}$ possible sparse vectors
  - Still of theoretical importance and may lead to practical consequences and insights (*e.g.* , bounds performance of tractable reconstruction methods, like $\ell_1 - norm$ optimization)

The CS "wonder"

- Reconstruction with ℓ_1-norm

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathcal{D}\mathbf{x}_k = \mathcal{D}\mathbf{x}$$

  where $\|\mathbf{x}\|_1 \triangleq \sum_i |x_i|$

- Prefect reconstruction $\hat{\mathbf{x}} = \mathbf{x}_k$ is possible with probability 1 at the cost of more required measurements w.r.t. ℓ_0-norm minimization
- However the ℓ_1-norm optimization problem is tractable with polynomial complexity
- We do not address ℓ_1-norm optimization in this talk

The CS "wonder"

- Reconstruction with ℓ_1-norm

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathcal{D}\mathbf{x}_k = \mathcal{D}\mathbf{x}$$

  where $\|\mathbf{x}\|_1 \triangleq \sum_i |x_i|$

- Prefect reconstruction $\hat{\mathbf{x}} = \mathbf{x}_k$ is possible with probability 1 at the cost of more required measurements w.r.t. ℓ_0-norm minimization
- However the ℓ_1-norm optimization problem is tractable with polynomial complexity
- We do not address ℓ_1-norm optimization in this talk

The CS "wonder"

- Reconstruction with ℓ_1-norm

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_1 \quad \text{subject to} \quad \boldsymbol{\mathcal{D}}\mathbf{x}_k = \boldsymbol{\mathcal{D}}\mathbf{x}$$

  where $\|\mathbf{x}\|_1 \triangleq \sum_i |x_i|$
- Prefect reconstruction $\hat{\mathbf{x}} = \mathbf{x}_k$ is possible with probability 1 at the cost of more required measurements w.r.t. ℓ_0-norm minimization
- However the ℓ_1-norm optimization problem is tractable with polynomial complexity
- We do not address ℓ_1-norm optimization in this talk

The CS "wonder"

- Reconstruction with ℓ_1-norm

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_1 \quad \text{subject to} \quad \boldsymbol{\mathcal{D}}\mathbf{x}_k = \boldsymbol{\mathcal{D}}\mathbf{x}$$

  where $\|\mathbf{x}\|_1 \triangleq \sum_i |x_i|$
- Prefect reconstruction $\hat{\mathbf{x}} = \mathbf{x}_k$ is possible with probability 1 at the cost of more required measurements w.r.t. ℓ_0-norm minimization
- However the ℓ_1-norm optimization problem is tractable with polynomial complexity
- We do not address ℓ_1-norm optimization in this talk

Noisy compressed sensing

- Underdetermined (overloaded) Gaussian vector channel

$$\mathbf{y} = \sqrt{\frac{\mathsf{snr}}{m}}\mathbf{D}\mathbf{x} + \mathbf{w}$$

Noisy compressed sensing

- Underdetermined (overloaded) Gaussian vector channel

$$\mathbf{y} = \sqrt{\frac{\mathsf{snr}}{m}}\mathbf{D}\mathbf{x} + \mathbf{w}$$

- Input $\mathbf{x}$ is κ_x-sparse

Noisy compressed sensing

- Underdetermined (overloaded) Gaussian vector channel

$$\mathbf{y} = \sqrt{\frac{\mathsf{snr}}{m}}\mathbf{D}\mathbf{x} + \mathbf{w}$$

- Input $\mathbf{x}$ is κ_x-sparse
- $\mathbb{E}\{x_i^2\}$ is finite

Noisy compressed sensing

- Underdetermined (overloaded) Gaussian vector channel

$$\mathbf{y} = \sqrt{\frac{\mathsf{snr}}{m}}\mathbf{D}\mathbf{x} + \mathbf{w}$$

- Input $\mathbf{x}$ is κ_x-sparse
- $\mathbb{E}\{x_i^2\}$ is finite
- Examples of overloaded Gaussian channels:

D	m	n
CDMA spreading matrix	processing gain	users
MIMO fading channel	rx. antennas	tx. antennas

Noisy compressed sensing via WGN sparse representation

- Let $\mathbf{z}_{k^*_\alpha} \in \mathbb{R}^n$ be a k^*_α-sparsest WGN representation

Noisy compressed sensing via WGN sparse representation

- Let $\mathbf{z}_{k_\alpha^*} \in \mathbb{R}^n$ be a k_α^*-sparsest WGN representation
- Thus for the dictionary/channel $\mathbf{D}$

$$\mathbf{y} = \sqrt{\frac{\mathsf{snr}}{m}}\mathbf{D}\mathbf{x} + \mathbf{w} = \sqrt{\frac{1}{m}}\mathbf{D}(\sqrt{\mathsf{snr}}\mathbf{x} + \sqrt{\alpha}\mathbf{z}_{k_\alpha^*}) = \sqrt{\frac{1}{m}}\mathbf{D}\mathbf{x}^*$$

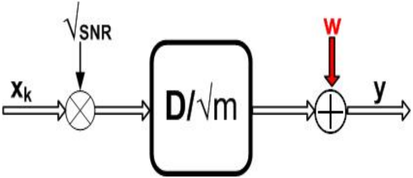
Noisy compressed sensing via WGN sparse representation

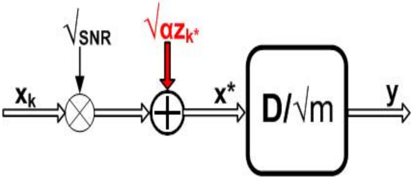
- Let $\mathbf{z}_{k_\alpha^*} \in \mathbb{R}^n$ be a k_α^*-sparsest WGN representation
- Thus for the dictionary/channel $\mathbf{D}$

$$\mathbf{y} = \sqrt{\frac{\text{snr}}{m}}\mathbf{D}\mathbf{x} + \mathbf{w} = \sqrt{\frac{1}{m}}\mathbf{D}(\sqrt{\text{snr}}\mathbf{x} + \sqrt{\alpha}\mathbf{z}_{k_\alpha^*}) = \sqrt{\frac{1}{m}}\mathbf{D}\mathbf{x}^*$$

- $\mathbf{x}^*$ is the sparsest representation of the observations $\mathbf{y}$ w.r.t. the dictionary/channel $\mathbf{D}$

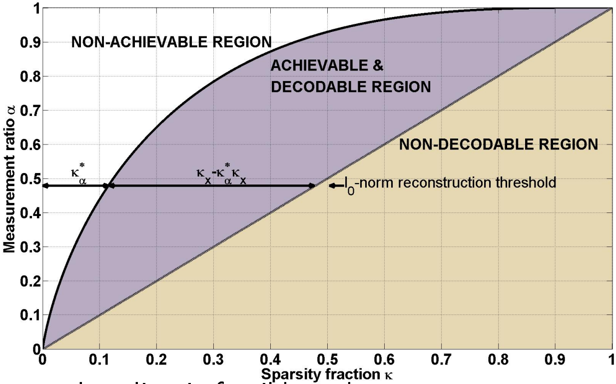
From noisy to "noiseless" channel





- The noisy channel with κ_x-sparse input is translated into a noiseless channel with $(\kappa_x + \kappa_\alpha^* - \kappa_x\kappa_\alpha^*)$-sparse input

ℓ_0-norm decoding of Gaussian vector channel

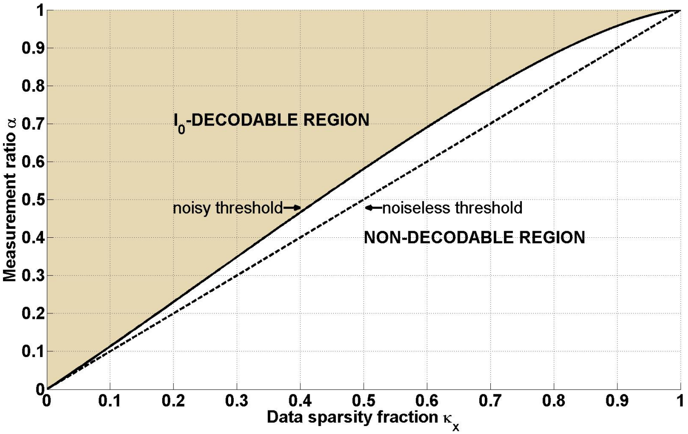


- ℓ_0-norm decoding is feasible as long as

$$\begin{aligned} \kappa_\alpha^* + \kappa_x - \kappa_\alpha^* \kappa_x &\leq \alpha \\ \Rightarrow \kappa_x &\leq \frac{\alpha - \kappa_\alpha^*}{1 - \kappa_\alpha^*} \end{aligned}$$

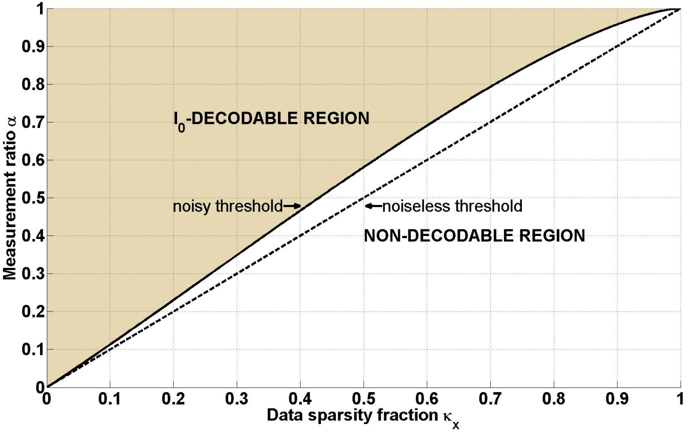
ℓ_0-norm decoding of Gaussian vector channel

$$\kappa_x \leq \frac{\alpha - \kappa_\alpha^*}{1 - \kappa_\alpha^*}$$



ℓ_0-norm decoding of Gaussian vector channel

$$\kappa_x \leq \frac{\alpha - \kappa_\alpha^*}{1 - \kappa_\alpha^*}$$



- Insensitive to snr: same decodable region for all SNR

ℓ_0-norm decoding of Gaussian vector channel

- The support Ω_0 of $\mathbf{x}^*$ is now known, thus the problem is well-posed and one could reconstruct $\mathbf{x}$ with Least-Squares method

$$\hat{\mathbf{x}}_{LS} = \begin{cases} \sqrt{\frac{m}{\text{snr}}}(\mathbf{D}_{\Omega_0}^T \mathbf{D}_{\Omega_0})^{-1} \mathbf{D}_{\Omega_0}^T \mathbf{y} & \text{on } \Omega_0 \\ 0 & \text{elsewhere} \end{cases}$$

ℓ_0-norm decoding of Gaussian vector channel

Proposition

Given a Gaussian vector channel with measurement ratio $\alpha \in (0,1)$ and arbitrary snr > 0, then in the large-system limit an ℓ_0-norm decoder results, with probability 1, in average mean-square error (MSE) per unknown

$$\frac{1}{n}\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 = \frac{\kappa_x + \kappa_\alpha^* - \kappa_x \kappa_\alpha^*}{snr},$$

as long as

$$\kappa_x \le \frac{\alpha - \kappa_\alpha^*}{1 - \kappa_\alpha^*}$$

for almost any $\mathbf{x}$. *Otherwise ℓ_0-reconstruction is impossible with overwhelming probability.*

ℓ_0-norm vs. oracle decoder

- In the absence of any other (prior) information, reconstruction with MSE proportional to the noise level, as happens for ℓ_0-norm decoding, is the best one can hope for

$$\text{MSE}_{\ell_0} = \frac{\kappa_x + \kappa^* - \kappa_x \kappa^*}{\text{snr}} \geq \frac{\kappa_x}{\text{snr}} = \text{MSE}_{\text{oracle}}$$

Universality

- For any orthonormal basis matrix $\mathbf{\Psi}$ (*e.g.*, DFT matrix) and a Gaussian dictionary $\mathbf{D}$, the matrix $\mathbf{D\Psi}$ will be also a Gaussian dictionary

Take-home message

- Introducing achievable and converse regions for sparse representation of WGN via Replica method
- The marginal distribution of such sparse representations is derived
- Introducing sharp threshold for ℓ_0-norm decoding in noisy compressed sensing
- The MSE of ℓ_0-norm decoder in underdetermined Gaussian vector channels is

$$\frac{\kappa_x + \kappa_\alpha^* - \kappa_x \kappa_\alpha^*}{\mathsf{snr}}$$

- Other applications (in, *e.g.*, data hiding, cryptography,...)?

Take-home message

- Introducing achievable and converse regions for sparse representation of WGN via Replica method
- The marginal distribution of such sparse representations is derived
- Introducing sharp threshold for ℓ_0-norm decoding in noisy compressed sensing
- The MSE of ℓ_0-norm decoder in underdetermined Gaussian vector channels is

$$\frac{\kappa_x + \kappa_\alpha^* - \kappa_x \kappa_\alpha^*}{\mathsf{snr}}$$

- Other applications (in, *e.g.*, data hiding, cryptography,...)?

Take-home message

- Introducing achievable and converse regions for sparse representation of WGN via Replica method
- The marginal distribution of such sparse representations is derived
- Introducing sharp threshold for ℓ_0-norm decoding in noisy compressed sensing
- The MSE of ℓ_0-norm decoder in underdetermined Gaussian vector channels is
$$\frac{\kappa_x + \kappa_\alpha^* - \kappa_x \kappa_\alpha^*}{\mathsf{snr}}$$
- Other applications (in, *e.g.*, data hiding, cryptography,...)?

Take-home message

- Introducing achievable and converse regions for sparse representation of WGN via Replica method
- The marginal distribution of such sparse representations is derived
- Introducing sharp threshold for ℓ_0-norm decoding in noisy compressed sensing
- The MSE of ℓ_0-norm decoder in underdetermined Gaussian vector channels is

$$\frac{\kappa_x + \kappa_\alpha^* - \kappa_x \kappa_\alpha^*}{\mathsf{snr}}$$

- Other applications (in, *e.g.*, data hiding, cryptography,...)?

Take-home message

- Introducing achievable and converse regions for sparse representation of WGN via Replica method
- The marginal distribution of such sparse representations is derived
- Introducing sharp threshold for ℓ_0-norm decoding in noisy compressed sensing
- The MSE of ℓ_0-norm decoder in underdetermined Gaussian vector channels is

$$\frac{\kappa_x + \kappa_\alpha^* - \kappa_x \kappa_\alpha^*}{\mathsf{snr}}$$

- Other applications (in, *e.g.*, data hiding, cryptography,...)?

- Questions?
- E-mail: oshental@qualcomm.com

THANK YOU!