What can be recovered under sparse adversarial corruption?

Vishal Halder, **Alexandre Reiffers-Masson**, Abdeldjalil Aïssa-El-Bey, Gugan Thoppe

Lab-STICC

IMT Atlantique

24 February 2026

Why care about sparse adversarial corruption?

In many systems, measurements can be corrupted by a small number of malicious or faulty sensors.

Examples.

- Network monitoring: a few compromised routers falsify traffic data.
- Power grids: attacked smart meters inject false readings.
- Distributed sensing: some sensors are hijacked.
- Medical imaging: sparse artifacts or adversarial perturbations.

Theoretical Model

Common structure.

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^\star + \boldsymbol{e}, \qquad \|\boldsymbol{e}\|_0 \leq q$$

Known	**Unknown**
Measurement matrix $\boldsymbol{A}$	Signal $\boldsymbol{x}^\star$
Observation $\boldsymbol{y}$	Sparse corruption $\boldsymbol{e}$ ($\|\boldsymbol{e}\|_0 \leq q$)

Classical question: Can we recover $\boldsymbol{x}^\star$ exactly and how?

But is it the right question?

Why classical recovery may fail

Even if only q measurements are corrupted:

- The adversary chooses the worst possible entries.
- Corruptions can mimic another valid signal.

There may exist two distinct signals $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$ such that

$$\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{e}_1 = \boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{e}_2, \quad \|\boldsymbol{e}_1\|_0, \|\boldsymbol{e}_2\|_0 \leq q.$$

Consequence: Exact recovery of $\boldsymbol{x}^\star$ is fundamentally impossible for most $\boldsymbol{A}$.

So the right question is not: "Can we recover everything?" but rather:

"What is guaranteed to survive?"

Why classical recovery may fail

Even if only q measurements are corrupted:

- The adversary chooses the worst possible entries.
- Corruptions can mimic another valid signal.

There may exist two distinct signals $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$ such that

$$\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{e}_1 = \boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{e}_2, \quad \|\boldsymbol{e}_1\|_0, \|\boldsymbol{e}_2\|_0 \leq q.$$

Consequence: Exact recovery of $\boldsymbol{x}^\star$ is fundamentally impossible for most $\boldsymbol{A}$.

So the right question is not: "Can we recover everything?" but rather:

Which components or linear combinations of $x^\star$ are invariant to any q corruptions?

From exact recovery to robust information

Main result: The recoverable information forms a linear subspace:

$$\mathcal{R} = \bigcap_{|T|=m-2q} \operatorname{rowspan}(\boldsymbol{A}_T),$$

where $\boldsymbol{A}_T$ denotes a row-deleted submatrix.

- $\mathcal{R}$ is the subspace common to the rowspans of every $2q$ row deleted submatrices.
- There can be nothing more that can be robustly recovered.

The robust projector

Define $\boldsymbol{U} = \mathrm{Proj}(\mathcal{R})$ as the *orthogonal projector* onto the robust subspace.

Main result: For any two feasible signals x_1, x_2 such that

$$\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{e}_1 = \boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{e}_2, \quad \|\boldsymbol{e}_1\|_0, \|\boldsymbol{e}_2\|_0 \leq q,$$

we necessarily have $\boldsymbol{U}\boldsymbol{x}_1 = \boldsymbol{U}\boldsymbol{x}_2$.

Interpretation.

- $\boldsymbol{U}\boldsymbol{x}^\star$ is the *maximally recoverable information*.
- $\ker(\boldsymbol{U})$ represents the fundamentally ambiguous directions.

A simple example: what *is* robustly recoverable?

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Let $q = 1$. Then, what about $\boldsymbol{x}^\star \in \mathbb{R}^5$ *can* be robustly recovered from $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^\star + \boldsymbol{e}$?

Answer.

$$\frac{1}{3}\sum_{i=1}^{3} x_i^\star, \qquad \frac{1}{2}\sum_{i=4}^{5} x_i^\star.$$

How do we formalize this for any $\boldsymbol{A}$?

A simple example: what *is* robustly recoverable?

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\mathrm{rowspan}(\boldsymbol{A}) = \mathrm{span}(\boldsymbol{u}, \boldsymbol{v}), \quad \boldsymbol{u} = \begin{bmatrix} 1\,1\,1\,0\,0 \end{bmatrix}^\top, \qquad \boldsymbol{v} = \begin{bmatrix} 0\,0\,0\,1\,1 \end{bmatrix}^\top.$$

Notice that deleting any $2q = 2$ rows from $\boldsymbol{A}$ preserves this as the rowspan, and

$$\frac{\boldsymbol{u}^\top \boldsymbol{x}^\star}{\|\boldsymbol{u}\|^2} = \frac{1}{3}\sum_{i=1}^{3} x_i^\star, \qquad \frac{\boldsymbol{v}^\top \boldsymbol{x}^\star}{\|\boldsymbol{v}\|^2} = \frac{1}{2}\sum_{i=4}^{5} x_i^\star.$$

Why $2q$?

Definition (Ambiguity sets)

For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and integer $q < m/2$, define the *ambiguity set* $S_q^{\boldsymbol{A}} := \{\boldsymbol{v} \in \mathbb{R}^n : \|\boldsymbol{A}\boldsymbol{v}\|_0 \leq 2q\}$.

Example

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}}_{\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{e}_1} = \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{\boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{e}_2} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 2 \end{bmatrix}}_{\boldsymbol{y}}$$

so, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are indistinguishable to the observer, and $\boldsymbol{x}_1 - \boldsymbol{x}_2 \in S_q^{\boldsymbol{A}}$, since $\|\boldsymbol{e}_1 - \boldsymbol{e}_2\|_0 \leq 2q$.

What does it mean to be *robust*?

Robust Functions: For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and integer $q < m/2$, a function $\mathscr{G} : \mathbb{R}^n \to \mathcal{Z}$ (arbitrary codomain) is said to be $(\boldsymbol{A}, q)$-*robust* if for every $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$, and q-sparse $\boldsymbol{e}_1, \boldsymbol{e}_2 \in \mathbb{R}^m$,

$$\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{e}_1 = \boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{e}_2 \implies \mathscr{G}(\boldsymbol{x}_1) = \mathscr{G}(\boldsymbol{x}_2).$$

Robust Solution Sets: Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $q < m/2$ be an integer. Then, for any $(\boldsymbol{A}, q)$-robust function $\mathscr{G}$ and any $\boldsymbol{x}^\star \in \mathbb{R}^n$, the $\mathscr{G}$-*robust solution set containing* $\boldsymbol{x}^\star$ is

$$X(\mathscr{G}, \boldsymbol{x}^\star) := \{\boldsymbol{x} \in \mathbb{R}^n : \mathscr{G}(\boldsymbol{x}) = \mathscr{G}(\boldsymbol{x}^\star)\}.$$

Clearly, $X(\mathscr{G}, \boldsymbol{x}^\star) = \mathscr{G}^{-1}(\mathscr{G}(\boldsymbol{x}^\star))$, the preimage of $\mathscr{G}(\boldsymbol{x}^\star)$ under $\mathscr{G}$.

Main Result

Theorem: Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and an integer $q < m/2$. If we define

$$\mathcal{R} = \bigcap_{\substack{T \subseteq [m], \\ |T| = m - 2q}} \operatorname{rowspan}(\boldsymbol{A}_T)$$

and $\boldsymbol{U}$ being the orthogonal projector onto $\mathcal{R}$. Then,

1. the $\boldsymbol{x} \to \boldsymbol{U}\boldsymbol{x}$ map is $(\boldsymbol{A}, q)$-robust.
2. For any $(\boldsymbol{A}, q)$-robust $\mathscr{G} : \mathbb{R}^n \to \mathcal{Z}$ and $\boldsymbol{x}^\star \in \mathbb{R}^n$,

$$\{\boldsymbol{x} \in \mathbb{R}^n : \mathscr{G}(\boldsymbol{x}) = \mathscr{G}(\boldsymbol{x}^\star)\} \supseteq \boldsymbol{x}^\star + \ker(\boldsymbol{U}),$$

with equality for $\mathscr{G}(\boldsymbol{x}) = \boldsymbol{U}\boldsymbol{x}$.

In words: the inclusion-wise minimal solution set that is robust to q adversarial corruptions is $\boldsymbol{x}^\star + \ker(\boldsymbol{U})$.

L0 decoder is *nearly* all you need

Theorem: Let $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^\star + \boldsymbol{e}$ with $\|\boldsymbol{e}\|_0 \leq q$, and let $\boldsymbol{U}$ be the matrix as defined in the previous theorem. Then every $\hat{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_0$ satisfies $\boldsymbol{U}\hat{\boldsymbol{x}} = \boldsymbol{U}\boldsymbol{x}^\star$.

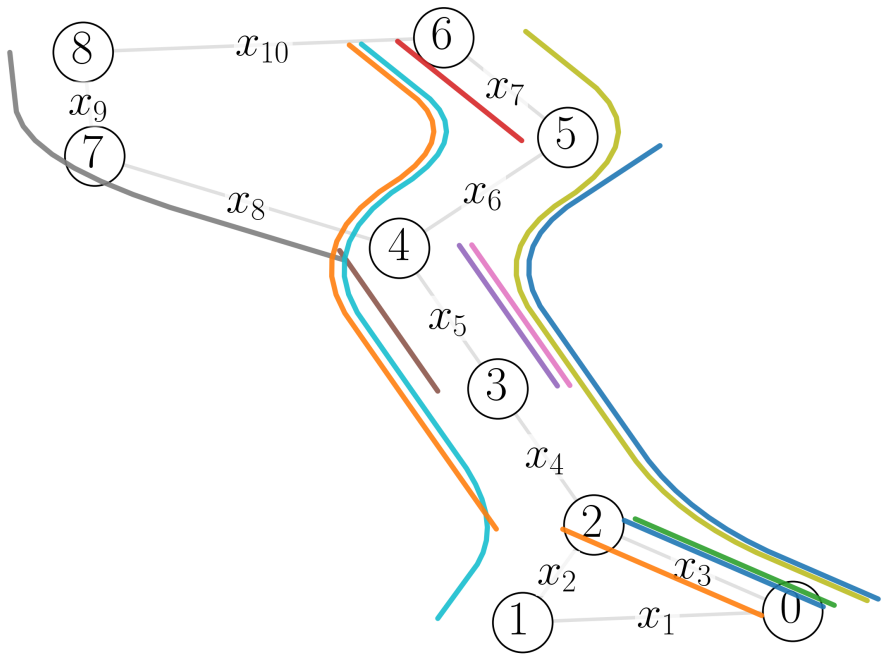
How to recover all recoverable information about $\boldsymbol{x}^\star$, for a given $\boldsymbol{A}, q$ and $\boldsymbol{y}$?

1. compute $\hat{\boldsymbol{x}}$ using the ℓ_0-decoder,
2. compute $\boldsymbol{U}$,
3. return $\hat{\boldsymbol{x}} + \ker(\boldsymbol{U})$.

Remark: There is an an algorithm to compute $\boldsymbol{U}$ for which the complexity is dominated by $\binom{m}{2q}$, making the algorithm exponential in m. Moreover we have proved that finding $\boldsymbol{U}$ is NP-Hard.

Stylized Application 1: Robust Network Tomography

Setup. Link metrics $x_1, \ldots, x_{10}$ are measured through *path sums*. Up to q path measurements are adversarially corrupted. We build the 0–1 path-link matrix $\boldsymbol{A}$ and compute the robust projector $\boldsymbol{U}$.



(a) Network and observed measurement paths

What we extract.

- $\mathrm{im}(\boldsymbol{U})$: robust *subspace* of link quantities
- basis vectors: robust *linear combinations*
- $\mathrm{diag}(\boldsymbol{U})$: per-link *robustness score*

Question. Which links (or combinations of links) are *provably invariant* to any q corrupted paths?

Stylized Application 1: Robust projector

$$\boldsymbol{U} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Geometric meaning. $\mathrm{im}(\boldsymbol{U})$ is the maximal subspace of link quantities that is invariant to any single corrupted path measurement.

Stylized Application 1: Robust projector

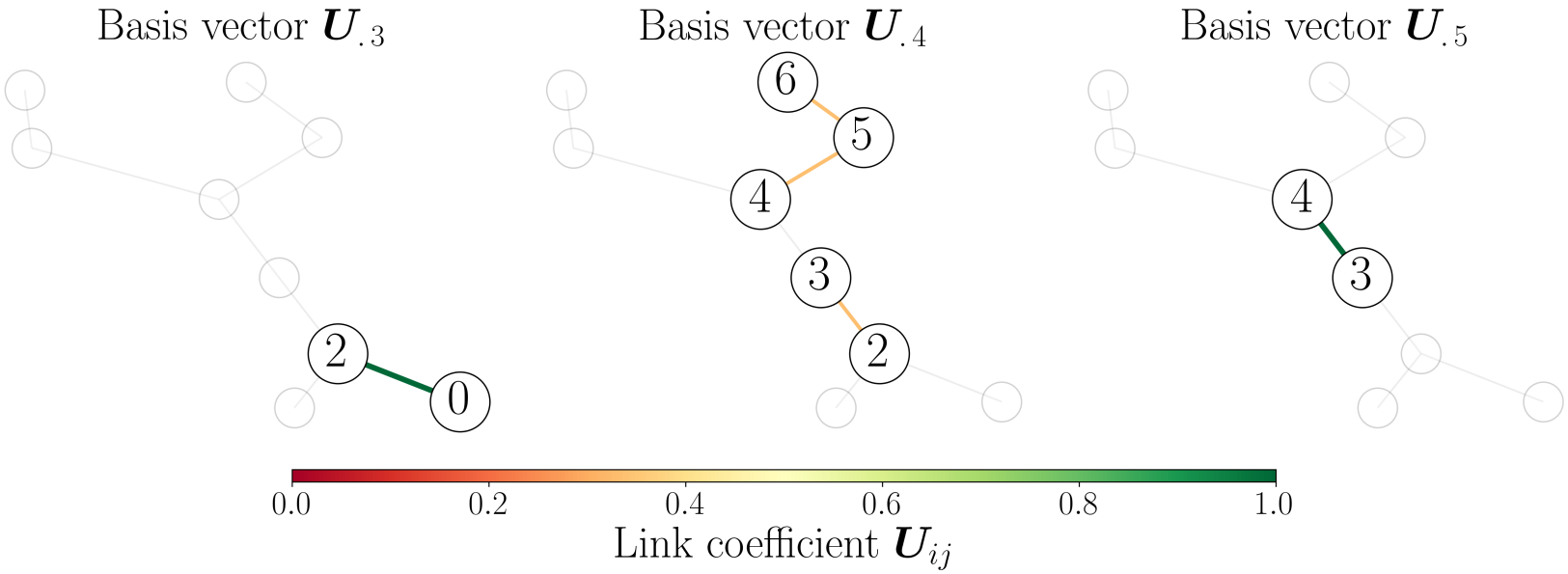


Figure: Basis vectors of $\mathrm{im}(\boldsymbol{U})$

Geometric meaning. $\mathrm{im}(\boldsymbol{U})$ is the maximal subspace of link quantities that is invariant to any single corrupted path measurement.

Stylized Application 1: Interpretation

What does the robust subspace tell us?

- $\mathrm{rank}(\boldsymbol{U}) = 3$: only a 3-dimensional subspace is robust.
- **Individually robust links:** x_3 and x_5.
- **Robust linear combination:** only $x_4 + x_6 + x_7$ is robust (not the links separately).

Interpretation. The projector $\boldsymbol{U}$ keeps exactly the link information that cannot be altered by any single corrupted path measurement.

Stylized Application 1: Robustness scores

Definition (Robustness score). For each link j, define $\text{score}(j) := U_{jj}$.

- $U_{jj} = 1 \quad \Rightarrow$ link j is individually robust.
- $U_{jj} = 0 \quad \Rightarrow$ link j is not robust.
- $0 < U_{jj} < 1 \quad \Rightarrow$ link j participates in a robust linear combination.

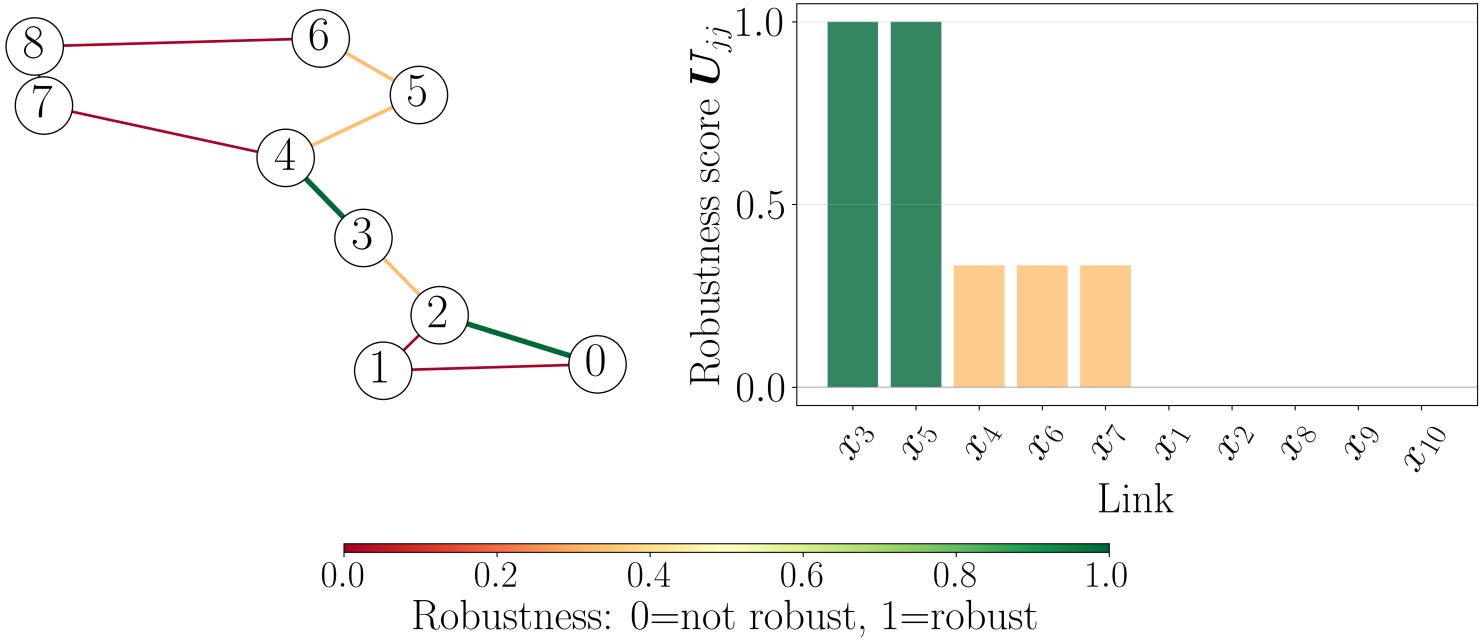


Figure: Visualization of $\text{diag}(\boldsymbol{U})$ as per-link robustness scores

Stylized Application 2: Oversampled DCT (Simulation 1)

Setup. Build $\boldsymbol{A} \in \mathbb{R}^{140\times 512}$ by sampling (with replacement) rows from a 512×512 DCT matrix. Corruption $\boldsymbol{e}$ is q-sparse with $q = 1$.

Orthonormal-row shortcut (counting rule). Distinct DCT rows are orthonormal, so robust directions correspond to DCT atoms that appear *many times* in $\boldsymbol{A}$.

$$\text{count(atom)} > 2q.$$

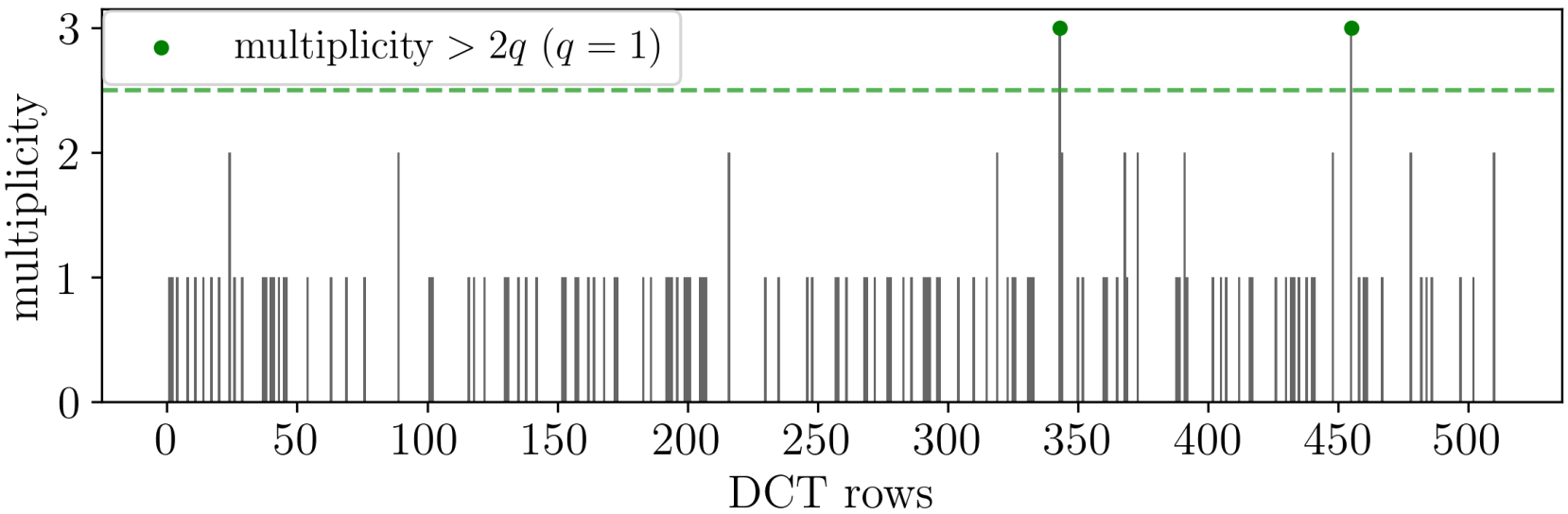
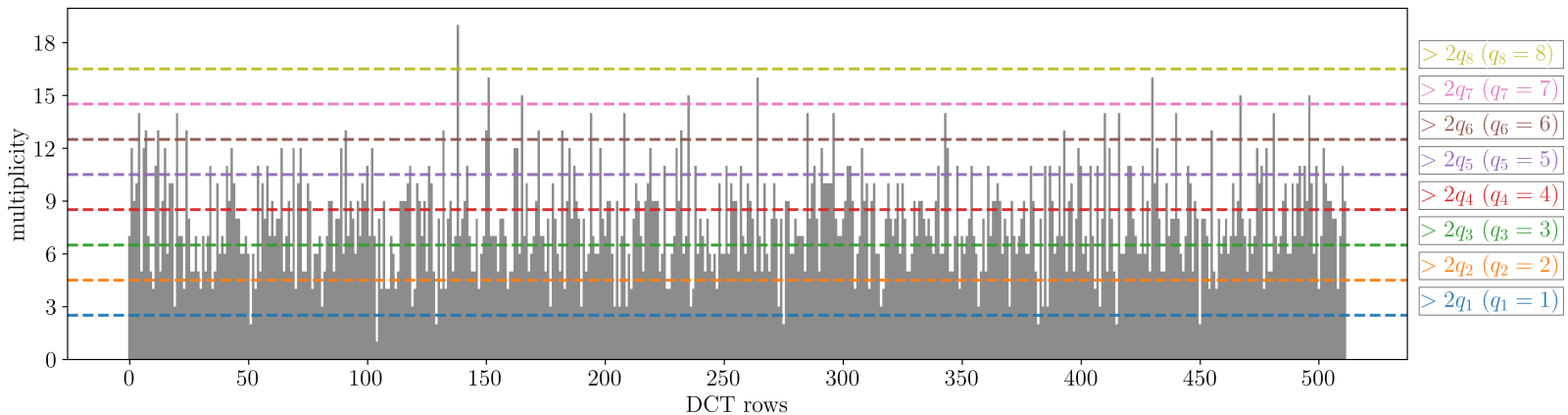


Figure: (a) Multiplicity of each sampled DCT atom, count > 2.

Oversampled DCT (Simulation 2)

Setup. Build $\boldsymbol{A} \in \mathbb{R}^{3600\times 512}$ by sampling DCT rows (with replacement). Vary $q \in \{1, \ldots, 8\}$ and count atoms satisfying

$$\text{count(atom)} > 2q.$$



(b) Robust atoms shrink as q increases (multiplicity threshold $2q$)

Oversampled DCT (Simulation 2) – observation & message

Observed phenomenon.

- As q increases, the number of robust DCT atoms drops sharply.
- At $q = 8$, only *one* atom remains robust.
- A slightly stronger adversary collapses the robust subspace to $\{0\}$.

Interpretation. Uniform sampling creates many *low-multiplicity* atoms; a sparse adversary can "erase" them by corrupting only a tiny fraction of measurements.

Takeaway. Even with large m (here 3600), certifiable robustness can be extremely fragile if the measurement design does not enforce redundancy.

Extensions: Nonlinear and Noisy Settings

1. Nonlinear measurements

$$\boldsymbol{y} = f(\boldsymbol{x}^\star) + \boldsymbol{e}$$

- What replaces row-space intersection?
- Robust invariant *manifold* instead of subspace?
- Local robustness via Jacobian analysis?

2. Sparse corruption + dense noise

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^\star + \boldsymbol{e} + \eta$$

- Robust subspace becomes approximate.
- Quantify degradation with noise level.

Extensions: ℓ_1-Based Robust Recovery

Instead of worst-case invariance, solve:

$$\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_1$$

Key questions

- When does ℓ_1 recovery implicitly extract $\boldsymbol{U}\boldsymbol{x}^\star$?
- Relation to nullspace / restricted nullspace conditions?
- Exact recovery vs partial (projected) recovery?

Goal

Bridge geometric robustness (projector $\boldsymbol{U}$) with convex optimization guarantees.

PhD Position: Online Machine Learning Algorithms for Optimal Cyber-Defense in Large Networks

Supervisors	Prof. Y. Hayel (Avignon University) Prof. A. Reiffers-Masson (IMT Atlantique)
Location	Avignon University (with visits to IMT Atlantique)
Keywords	Machine Learning, Optimization, Cybersecurity

Research Objectives

- Online estimation of spectral properties of large unknown graphs
- Controlled random walks + reinforcement learning for graph exploration
- Cyber-defense via node removal to minimize attack hitting probabilities

Contact: yezekael.hayel@univ-avignon.fr — alexandre.reiffers-masson@imt-atlantique.fr

Papers on adversarial learning

- Halder, V., Reiffers-Masson, A., Aïssa-El-Bey, A., & Thoppe, G. (2025). What Can Be Recovered Under Sparse Adversarial Corruption? Assumption-Free Theory for Linear Measurements. arXiv preprint arXiv:2510.24215.
- Reiffers-Masson, A., & Amigo, I. (2023). *Online Multi-Agent Decentralized Byzantine-robust Gradient Estimation.* ACM SIGMETRICS Performance Evaluation Review, 50(4), 38-40. Application to Multi-Agent Gradient Estimation.
- Ganesh, S., Reiffers-Masson, A., & Thoppe, G. (2023, December). *Online learning with adversaries: A differential-inclusion analysis.* In 2023 62nd IEEE Conference on Decision and Control (CDC) (pp. 1288-1293). IEEE. Almost sure convergence of the two-time scale algorithm.
- Thoppe, G., Dhankshiru, M., Roy, N., Reiffers-Masson, A., Naman, N., & Azor, A. (2024). *Adversary-Resilient Distributed Estimation using Intermittent and Heterogeneous Data with Application to Network Tomography.* Finite time convergence rate and application to network tomography.