Nick Vannieuwenhoven

KU LEUVEN

with Breiding, Gesmundo, and Michałek

Algebraic compressed sensing (of secant varieties)


There Will Be
Algebraic Geometry

Overview

Overview

Compressed sensing

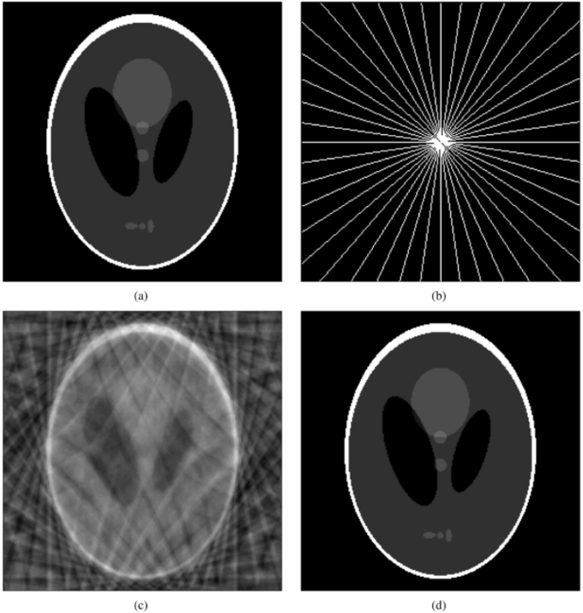
The **Whittaker–Shanon sampling theorem** was believed to be a fundamental boundary for exact reconstruction of functions in signal processing from finite samples, until about 2005.

Around that time, Donoho[1] and Candès, Romberg, and Tao[2] made the crucial observation that if the signals admit other structures than band-limitedness, reconstruction from **fewer samples than the Nyquist frequency** may be possible.

Donoho coined the term **compressed sensing** for this circle of ideas. (Alternative names include *compressive sensing* and *compressive sampling*).

[1] *Compressed sensing*, IEEE Trans. Inform. Theory, 2006.

[2] *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 2006



Candès, Romberg, and Tao gave a stunning illustration of exact reconstruction from a number of samples of about 2% of the Nyquist frequency.

The illustration is reproduced on the left from their IEEE Trans. Inform. Theory paper.

Sparse compressed sensing

Initially compressed sensing focused on the structure of k-**sparse vectors.** Recall that a vector $x \in \mathbb{k}^n$ is k-sparse if

$$\|x\|_0 := \sharp\{x_i \neq 0\} = k.$$

k-sparse compressed sensing

Assume that $x \in \mathbb{k}^n$ is a k-sparse vector. Given

1. a **linear sampling map** $\mu : \mathbb{k}^n \to \mathbb{k}^s$,
2. a **compressed sensing** $y = \mu(x)$,

then solve

$$\min_{x \in \mathbb{k}^n} \|x\|_0 \quad \text{s.t. } y = \mu(x)$$

Sparse compressed sensing

Initially compressed sensing focused on the structure of k-**sparse vectors.** Recall that a vector $x \in \mathbb{k}^n$ is k-sparse if

$$\|x\|_0 := \sharp\{x_i \neq 0\} = k.$$

k-sparse compressed sensing

Assume that $x \in \mathbb{k}^n$ is a k-sparse vector. Given

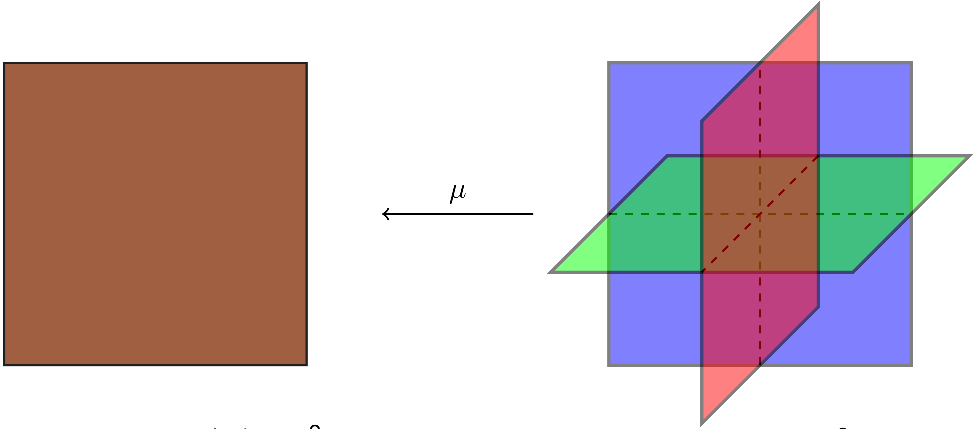
1. a **linear sampling map** $\mu : \mathbb{k}^n \to \mathbb{k}^s$,
2. a **compressed sensing** $y = \mu(x)$,

then solve

$$\min_{x \in \mathbb{k}^n} \|x\|_0 \quad \text{s.t. } y = \mu(x)$$

In practice, the equalities $=$ would be approximate equalities $\approx$, and the recovery problem becomes an **approximation problem**, to be solved with optimization algorithms.

Geometrically, for 2-sparse vectors in 3-space, we have the following setup:

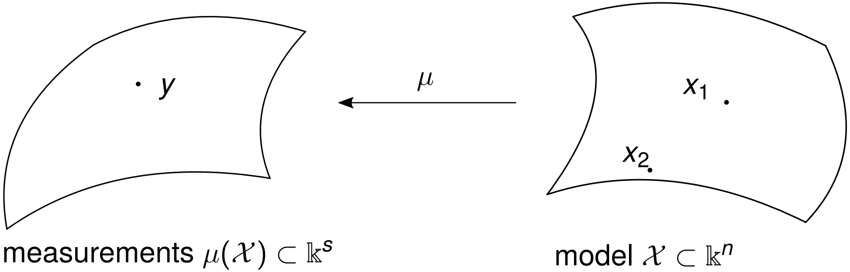


measurements $\mu(\mathcal{X}) \subset \mathbb{k}^2$

model $\mathcal{X} \subset \mathbb{k}^3$

General compressed sensing

Compressed sensing ideas were vastly extended by Baraniuk and Wakin[3] by relaxing the k-sparsity requirement (i.e., union of subspaces) to arbitrary **smooth manifolds**:



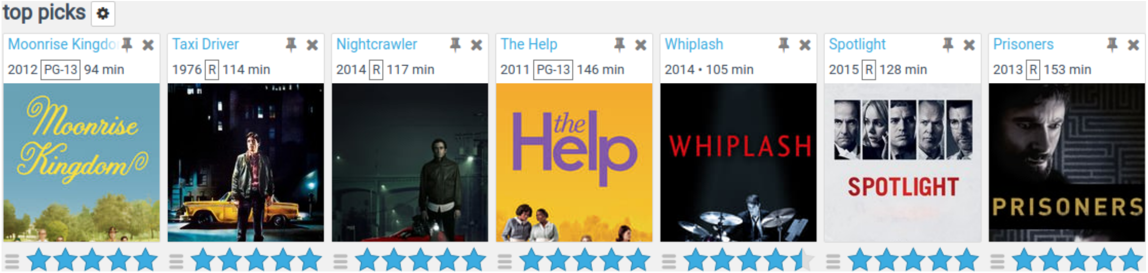
where $\mu : \mathbb{k}^n \to \mathbb{k}^s$ is still a linear map, defined on the whole ambient space.

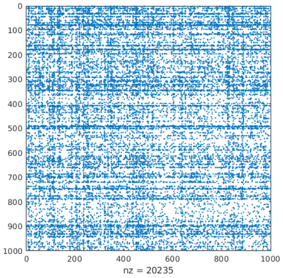
[3] *Random projections of smooth manifolds*, Found. Comput. Math., 2009.

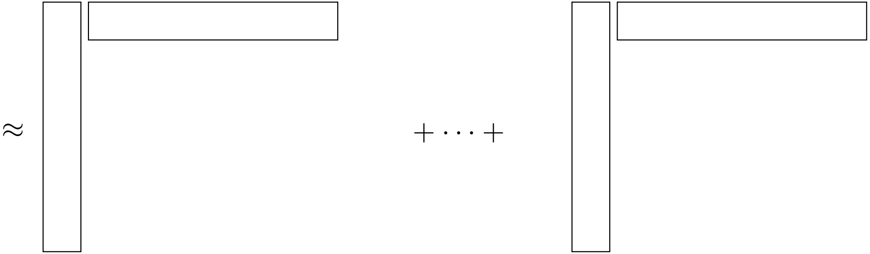
The general formulation of compressed sensing includes many important special cases:

1. k-sparse compressed sensing ($\mathcal{X}$ is a union of subspaces),
2. low-rank **matrix completion and recovery** ($\mathcal{X}$ is the manifold of rank-r matrices),
3. low-rank **tensor completion and recovery** ($\mathcal{X}$ is a low-rank tensor manifold),
4. recovery of distributions from incomplete moment sequences in **algebraic statistics.**

Example: matrix completion







Overview

How many samples ensure that a problem instance is well posed?

An compressed sensing problem is **well-posed** in the sense of Hadamard at $y \in \mathbb{k}^s$ if the following conditions simultaneously hold:[4]

(E) **Existence:**
There exists an $x \in \mathcal{X}$ with $y = \mu(x)$;

(I) **Identifiability:**
There is an open neighborhood U of y in $\mu(\mathcal{X})$ such that, for every $y' \in U$, there is a unique $x' \in \mathcal{X}$ with $y' = \mu(x')$;

(C) **Continuity:**
The solution $x' = \mu^{-1}(y')$ is continuous as a function of $y' \in U$.

[4] Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, Springer, 2011.

The null space property

Since μ is a linear map, the fiber of μ at $\mu(x)$ is a single point if and only if

$$\mu(x) - \mu(y) = 0 = \mu(x - y) \quad \text{implies that} \quad x = y.$$

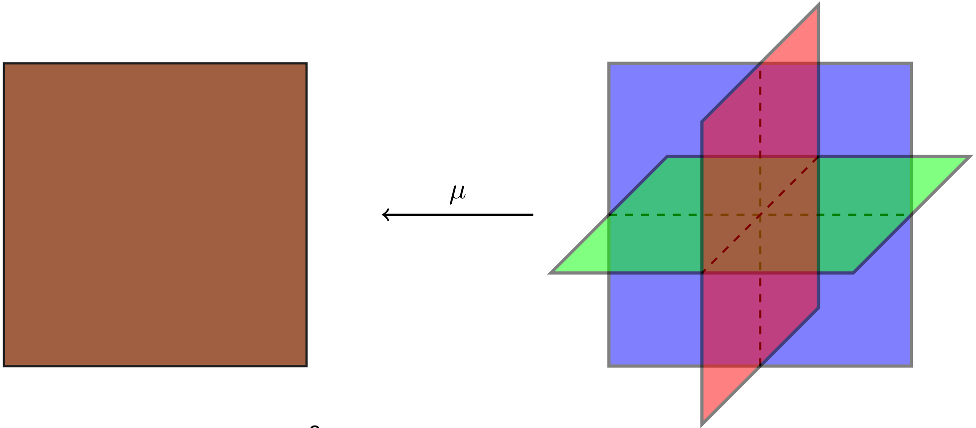
Hence, the compressed sensing problem has a unique solution if and only if the kernel of μ intersects the **set of differences**

$$\Delta(\mathcal{X}) = \{x - y \mid x, y \in \mathcal{X}\}$$

only trivially. (Or, alternatively, $x + \ker(\mu) \cap \mathcal{X} = \{x\}$ for all $x \in \mathcal{X}$.)

This is the essence of the **null space property** in compressed sensing. Verifying it in a given setup is considered a hard problem.

Thus, the compressed sensing problem below is well posed at generic points for a generic linear map $\mu : \mathbb{k}^3 \to \mathbb{k}^2$.



measurements $\mu(\mathcal{X}) \subset \mathbb{k}^2$

model $\mathcal{X} \subset \mathbb{k}^3$

The restricted isometry property

The most popular approach for establishing well posedness is the **restricted isometry property** (RIP).[5] A map $\mu : \mathbb{k}^n \supset \mathcal{X} \to \mathbb{k}^s$ satisfies the ϵ-RIP if

$$1 - \epsilon \leqslant \frac{\|\mu(x) - \mu(y)\|}{\|x - y\|} \leqslant 1 + \epsilon \quad \text{for all } x, y \in \mathcal{X}.$$

This states that μ must be an **almost isometric embedding** of $\mathcal{X}$ into $\mathbb{k}^s$.

If μ satisfies this property, then it is clearly injective, so a unique solution to the compressed sensing problem exists for all inputs.

[5] Candès, Tao, *Decoding by linear programming*, IEEE Trans. Inform. Process., 2005.

Note that the ϵ-RIP of μ implies that it is a **bi-Lipschitz embedding**, studied for general metric space since the late 1970's by Luukkainen and Väisälä,[6] Assouad,[7] and Johnson and Lindenstrauss.[8]

Recall that a map $\mu : \mathcal{X} \to \mathcal{Y}$ between metric spaces $(\mathcal{X}, \mathrm{dist}_{\mathcal{X}})$ and $(\mathcal{Y}, \mathrm{dist}_{\mathcal{Y}})$ is a bi-Lipschitz embedding if

$$0 < c \leqslant \frac{\mathrm{dist}_{\mathcal{Y}}(\mu(x), \mu(y))}{\mathrm{dist}_{\mathcal{X}}(x, y)} \leqslant C < \infty \quad \text{for all } x, y \in \mathcal{X}.$$

A bi-Lipschitz embedding has an inverse map μ^{-1} that is also bi-Lipschitz with upper and lower constants C^{-1} and c^{-1}.

[6] *Elements of Lipschitz topology*, Ann. Acad. Sci. Fenn. Math., 1977.
[7] *Plongements lipschitzien dans $\mathbb{R}^n$*, Bull. Soc. Math. France, 1983.
[8] *Extensions of Lipschitz mappings into a Hilbert space*, Contemp. Math., 1984.

For bi-Lipschitz embeddings, we find that the three well posedness conditions hold **at every point** of $\mu(\mathcal{X})$:

- $\mu^{-1} : \mu(\mathcal{X}) \to \mathcal{X}$ **exists**,
- μ^{-1} is a **function**, and
- μ^{-1} is **Lipschitz continuous**.

For bi-Lipschitz embeddings, we find that the three well posedness conditions hold **at every point** of $\mu(\mathcal{X})$:

- $\mu^{-1} : \mu(\mathcal{X}) \to \mathcal{X}$ **exists**,
- μ^{-1} is a **function**, and
- μ^{-1} is **Lipschitz continuous**.

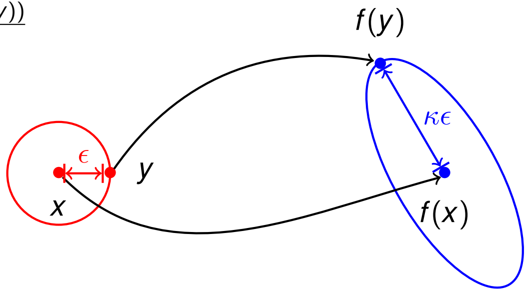
Analyzing the above properties implied by the RIP, we see this is

1. much more than what Hadamard asked,
2. restrictive from the viewpoint of **numerical analysis**, and
3. restrictive from the viewpoint of **algebraic geometry**.

What a numerical analyst may find curious

The **condition number** $\kappa[f](x)$ of a map $f : \mathcal{X} \to \mathcal{Y}$ between metric spaces $(\mathcal{X}, \mathrm{dist}_{\mathcal{X}})$ and $(\mathcal{Y}, \mathrm{dist}_{\mathcal{Y}})$ at a point $x \in \mathcal{X}$ is[9]

$$\kappa[f](x) = \lim_{\epsilon \to 0} \sup_{\substack{y \in \mathcal{X},\\ \mathrm{dist}_{\mathcal{X}}(x,y) \leqslant \epsilon}} \frac{\mathrm{dist}_{\mathcal{Y}}(f(x), f(y))}{\mathrm{dist}_{\mathcal{X}}(x,y)}$$



This number measures the **sensitivity** of f to infinitesimal perturbations. It can be regarded as a measure of **numerical hardness** of computing f.

[9] Rice, *A theory of condition*, SIAM J. Numer. Anal., 1966.

The global Lipschitzness of μ^{-1}, guaranteed by the RIP implies that the condition number of the reconstruction problem μ^{-1} is upper **bounded by a constant**!

Proposition

Let μ be a bi-Lipschitz mapping with constants c, C, then for all $y \in \mu(\mathcal{X})$, we have

$$0 < C^{-1} \leqslant \kappa[\mu^{-1}](y) \leqslant c^{-1} < \infty$$

This is strange from the viewpoint of numerical analysis.

All interesting computational problems f have a locus where $\kappa[f](x) \to \infty$, like solving

1. linear systems,
2. eigenproblems,
3. polynomial systems,
4. tensor rank decomposition,
5. etc.

For secant varieties, for example, there is a natural associated computational problem:

$$\pi^{-1} : \sigma_r(\mathcal{X}) \to \mathcal{X}^{[r]}, \quad x_1 + \cdots + x_r \mapsto \{x_1, \ldots, x_r\}$$

(where this is well defined; see Breiding and Vannieuwenhoven[10] for details).

[10] *The condition number of join decompositions*, SIAM J. Matrix Anal. Appl., 2018.

[11] *On the Terracini locus of projective varieties*, Milan J. Math., 2021.

[12] *Terracini locus for three points on a Segre variety*, arXiv, 2020.

For secant varieties, for example, there is a natural associated computational problem:

$$\pi^{-1} : \sigma_r(\mathcal{X}) \to \mathcal{X}^{[r]}, \quad x_1 + \cdots + x_r \mapsto \{x_1, \ldots, x_r\}$$

(where this is well defined; see Breiding and Vannieuwenhoven[10] for details).
We showed that

$$\kappa[\pi^{-1}](x_1 + \cdots + x_r) = \mathrm{dist}^{-1}((\mathrm{T}_{x_1}\mathcal{X}, \ldots, \mathrm{T}_{x_r}\mathcal{X}), \Sigma_{\mathrm{Gr}})$$

where Σ_{Gr} is what we called the ill-posed locus:

$$\Sigma_{\mathrm{Gr}} = \{(W_1, \ldots, W_r) \in \mathrm{Gr}(N, n) \times \cdots \times \mathrm{Gr}(N, n) \mid \dim(W_1 + \cdots + W_r) < rn\}.$$

[10] *The condition number of join decompositions*, SIAM J. Matrix Anal. Appl., 2018.
[11] *On the Terracini locus of projective varieties*, Milan J. Math., 2021.
[12] *Terracini locus for three points on a Segre variety*, arXiv, 2020.

For secant varieties, for example, there is a natural associated computational problem:

$$\pi^{-1} : \sigma_r(\mathcal{X}) \to \mathcal{X}^{[r]}, \quad x_1 + \cdots + x_r \mapsto \{x_1, \ldots, x_r\}$$

(where this is well defined; see Breiding and Vannieuwenhoven[10] for details).
We showed that

$$\kappa[\pi^{-1}](x_1 + \cdots + x_r) = \mathrm{dist}^{-1}((\mathrm{T}_{x_1}\mathcal{X}, \ldots, \mathrm{T}_{x_r}\mathcal{X}), \Sigma_{\mathrm{Gr}})$$

where Σ_{Gr} is what we called the ill-posed locus:

$$\Sigma_{\mathrm{Gr}} = \{(W_1, \ldots, W_r) \in \mathrm{Gr}(N, n) \times \cdots \times \mathrm{Gr}(N, n) \mid \dim(W_1 + \cdots + W_r) < rn\}.$$

This locus is nonempty, and is essentially the **Terracini locus**; see Breiding and Vannieuwenhoven,[10] Ballico and Chiantini,[11] and Ballico, Bernardi, Santarsiero.[12]

[10] *The condition number of join decompositions*, SIAM J. Matrix Anal. Appl., 2018.
[11] *On the Terracini locus of projective varieties*, Milan J. Math., 2021.
[12] *Terracini locus for three points on a Segre variety*, arXiv, 2020.

What a geometer may find curious

Proving the RIP of a fixed map μ may not be simpler than proving the null space property.

The main idea of Candès and Tao[13] was proving the RIP holds with high probability for **random matrices** sampled from a **Euclidean-open subset** of linear maps.

[13] *Decoding by linear programming*, IEEE Trans. Inform. Process., 2005.

A typical result, found in Foucart and Rauhut,[14] is as follows:

Theorem 9.2. *Let* $\mathbf{A}$ *be an* $m \times N$ *subgaussian random matrix. Then there exists a constant* $C > 0$ *(depending only on the subgaussian parameters* β, κ*) such that the restricted isometry constant of* $\frac{1}{\sqrt{m}}\mathbf{A}$ *satisfies* $\delta_s \leq \delta$ *with probability at least* $1 - \varepsilon$ *provided*

$$m \geq C\delta^{-2}\left(s \ln(eN/s) + \ln(2\varepsilon^{-1})\right). \tag{9.3}$$

Setting $\varepsilon = 2\exp(-\delta^2 m/(2C))$ yields the condition

$$m \geq 2C\delta^{-2} s \ln(eN/s),$$

which guarantees that $\delta_s \leq \delta$ with probability at least $1 - 2\exp\left(-\delta^2 m/(2C)\right)$. This is the statement often found in the literature.

[14] *A Mathematical Introduction to Compressive Sensing*, Birkhäuser, 2013.

This is strange from the viewpoint of geometry as well.

Specifically, if we assume our structured space $\mathcal{X} \subset \mathbb{k}^n$ is a smooth algebraic variety, the appearance of the **log-factor** in

$$s \geqslant \log(n) \cdot \dim \mathcal{X}$$

seems rather excessive, even if you need to restrict to linear bi-Lipschitz embeddings.

There are several global embedding theorems where $\mathcal{X}$ is a smooth variety suggesting you could do without this log factor:

1. **Whitney's smooth embedding theorem**[15] into $s = 2\dim\mathcal{X}$,
2. **Nash's isometric embedding theorem**[16] into $s \leqslant \frac{1}{2}\dim\mathcal{X}(3\dim\mathcal{X}+11)$.
3. **Birbrair, Fernandes, and Jelonek's bi-Lipschitz embedding theorem**[17] into $s = 2\dim\mathcal{X}+1$.

[15] *Elementary structure of real algebraic varieties*, Ann. Math., 1957.
[16] *The imbedding problem for Riemannian manifolds*, Ann. Math., 1956.
[17] *On the extension of bi-Lipschitz mappings*, Sel. Math. New Ser., 2021.

Overview

Algebraic compressed sensing

Breiding, Gesmundo, Michałek, and Vannieuwenhoven[18] propose to analyse **algebraic** compressed sensing problems:

$$\mu : \mathbb{k}^n \supset \mathcal{X} \to \mathbb{k}^s$$

where

- $\mathcal{X}$ is a **(semi-)algebraic (quasi-)variety** over $\mathbb{k} = \mathbb{R}$ or $\mathbb{C}$, and
- $\mu : \mathbb{k}^n \to \mathbb{k}^s$ is a linear map.

[18] *Algebraic compressed sensing*, arXiv, 2021.

Essentially, we translated most of the known results from algebraic geometry on these questions into the following result for quasi-varieties.

Theorem (The hitchhicker's guide to generic linear projections of varieties)

Let $\mathcal{X}$ be an ***irreducible variety*** *of dimension d over $\mathbb{k} = \mathbb{C}$ or $\mathbb{R}$, and let μ be a* ***generic linear map*** *$\mu : \mathcal{X} \to \mathbb{k}^s$. Then:*

(E) *the fiber $\mu^{-1}(y) \neq \emptyset$ at $y \in \mathbb{k}^s$ if and only if a particular system of polynomial equations has a solution.*

(R) *if $s \geqslant d$, the fiber $\mu^{-1}(\mu(x))$ of a generic point $x \in \mathcal{X}$ is finite;*

(I) *if $s \geqslant d + 1$, the fiber $\mu^{-1}(\mu(x))$ of a generic point $x \in \mathcal{X}$ is equal to $\{x\}$;*

(C) *if $s \geqslant d + 1$, the fiber $\mu^{-1}(\mu(x))$ of a generic point $x \in \mathcal{X}$ is C^∞ continuous.*

For completeness, the questions we considered were the following.

Existence:
Q1: Under what conditions do there exist solutions $y = \mu(x)$ for a given point $y \in \mathbb{k}^s$?

For completeness, the questions we considered were the following.

Recoverabilty:

Q2: Under what conditions is a generic linear map $\mu : \mathcal{X} \to \mathbb{k}^s$ **everywhere recoverable**?

Q3: Under what conditions is $\mu : \mathcal{X} \to \mathbb{k}^s$ defined by s generic (non-homogeneous) polynomials of degree p everywhere recoverable?

Q4: Under what conditions does there exist a **coordinate projection** $\hat{\mu} : \mathbb{k}^n \to \mathbb{k}^s$ that is generically recoverable when restricted to $\mathcal{X}$?

Q5: Given a **specific polynomial map** $\mu : \mathcal{X} \to \mathbb{k}^s$, how can one determine whether μ is generically recoverable?

Q6: If $\mathcal{X}$ is the image of a set of known polynomial equations, how can one **determine a subset of $s \geqslant d$ coordinates** such that the coordinate projection on these s coordinates is generically recoverable?

For completeness, the questions we considered were the following.

Identifiability:

Q7: Under what conditions is a generic linear map $\mu : \mathcal{X} \to \mathbb{k}^s$ generically identifiable?

Q8: Under what conditions is a map $\mu : \mathcal{X} \to \mathbb{k}^s$ defined by s generic polynomials of degree p generically identifiable?

Q9: Under which conditions is a generic linear map $\mu : \mathcal{X} \to \mathbb{k}^s$ **everywhere identifiable**?

Q10: Under what conditions does there exist a Euclidean open set of everywhere identifiable linear maps $\mu : \mathcal{X} \to \mathbb{R}^s$?

For completeness, the questions we considered were the following.

Continuity:

Q11: Under what conditions does a generic linear map $\mu : \mathcal{X} \to \mathbb{k}^s$ have the property that there exists a Zariski open manifold $\mathcal{X}' \subseteq \mathcal{X}$ such that $\mu(\mathcal{X}')$ is a smoothly embedded submanifold of $\mathbb{k}^s$ and $\mu|_{\mathcal{X}'}$ is a **local diffeomorphism** onto its image at every point $x \in \mathcal{X}'$?

Q12: Under what conditions does a generic linear map $\mu : \mathcal{X} \to \mathbb{k}^s$ have the property that there exists a Zariski open manifold $\mathcal{X}' \subset \mathcal{X}$ such that $\mu(\mathcal{X}')$ is a smoothly embedded submanifold of $\mathbb{k}^s$ and $\mu|_{\mathcal{X}'}$ is a **global diffeomorphism** onto its image?

Q13: How does one **compute the condition number** of inverting the projection at a smooth point $x \in \mathcal{X}$?

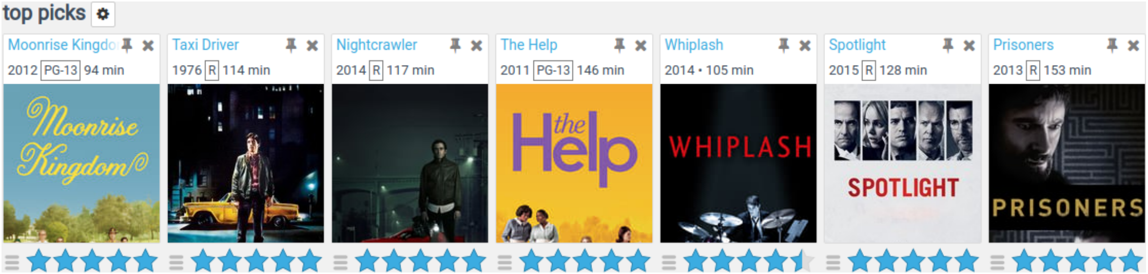
Overview

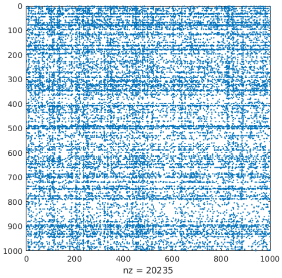
Coordinate projections

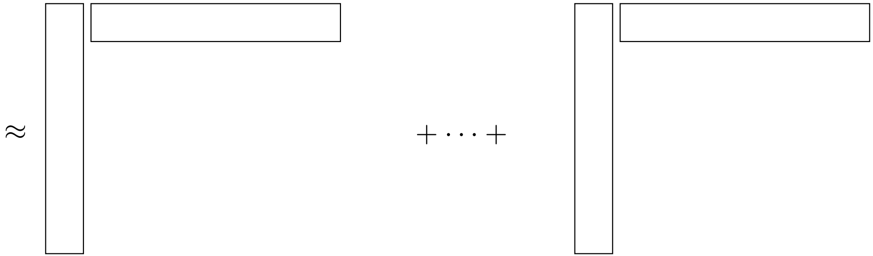
Other than randomly sampled linear maps (from subgaussian ensembles), one of the most important type of projections in applications are **coordinate projections.**

They arise in **matrix and tensor completion** problems, where not all data is (or can be) observed. The goal is to **predict the missing data** by making assumptions about the model from which the data originates.

Example: matrix completion







Which coordinates should we collect?

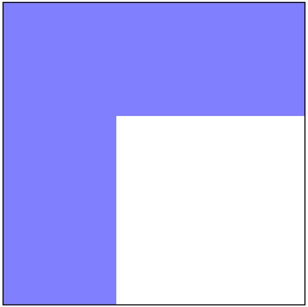
The central question surrounding coordinate projections is the following.

Which coordinates should we collect?

Let $\mathcal{X} \subset \mathbb{k}^n$ be an algebraic subvariety over a field of characteristic zero. Which $\dim \mathcal{X}$ $(+1)$ coordinates should we project to so that a generic point in the image is recoverable (identifiable)?

Example: low-rank matrices via cross approximation

Recall that $\dim \sigma_r(\mathrm{Seg}(\mathbb{k}^m \times \mathbb{k}^n)) = r(m+n-r)$. If we select any "cross" of r rows and r columns of $A \in \sigma_r(\mathrm{Seg}(\mathbb{k}^m \times \mathbb{k}^n)) \subset \mathbb{k}^{m \times n}$, i.e.,



then we can recover all coordinates of A from this minimal coordinate projection.

To the best of my knowledge, the question of **which coordinates** should be collected (and a **corresponding algorithm** for the reconstruction) for a minimal number of coordinate projections is **open** for:

1. the tensor rank decomposition,
2. Waring decomposition,
3. partially symmetric tensor decompositions,
4. (smooth/structured) block term decompositions,
5. Tucker decomposition,[19]
6. tensor trains decomposition,[20]
7. hierarchical Tucker decomposition[21]

---

[19] Reconstruction from a small but nonminimal number of coordinate projections in Caiafa and Cichocki, *Generalizing the column-row matrix decomposition to multiway arrays*, Linear Algebra Appl., 2010

[20] Cross approximation comes close to the minimal, see Oseledets and Tyrtyshnikov, *TT-cross approximation for multidimensional arrays*, Linear Algebra Appl., 2010.

[21] The hierarchical cross approximation scheme requires a low but nonminimal number of coordinates; see Ballani, Grasedyck, and Kluge, *Black-box approximation of tensors in hierarchical Tucker format*, Linear Algebra Appl., 2013.

More intrigue

There is a series of papers by Ashraphijuo[22] and one paper (plus a correction) by Pimentel-Alarcon that **claim to solve** the foregoing problem for, among others,

1. low-rank matrices,
2. tensor rank decomposition,
3. Tucker decomposition,
4. tensor trains decomposition.

However, the **proofs in these papers contain gaps.** The main one being, in algebraic parlance, that they use the false statement:

> *"polynomials that are algebraically independent form a regular sequence;"*

see Tsakiris[23] for more details.

---

[22] In J. Machine Learn. Res. 2×, IEEE Trans. Inform. Theory, IEEE Signal Process. Lett., Ann. Math. Artificial Intell., Pattern Recognition

[23] *Low-rank matrix completion theory via Plucker coordinates*, arXiv, 2021.

Overview

Conclusions

There are many interesting open questions on **coordinate projections of secant varieties** relevant to tensor completion problems!

Conclusions

There are many interesting open questions on **coordinate projections of secant varieties** relevant to tensor completion problems!

Thanks for your attention!
Dziękuję za uwagę!