Self Supervised Learning Methods for Imaging

Part 5: Identification Theory

Mike Davies, University of Edinburgh

Julián Tachella, CNRS, École Normale Supérieure de Lyon

Mathematical problems

1. **Signal Recovery:** Given the signal model $p_{\boldsymbol{x}}$, is there a unique $\boldsymbol{x}$ for $\boldsymbol{y} = A\boldsymbol{x}$

2. **Model Identification:** Can we *uniquely* identify the distribution $p_{\boldsymbol{x}}$ from the measurement distribution $p_{\boldsymbol{y}}$?

- All possible pairs of answers possible (eg. no signal recovery but model identification possible)

- Signal recovery has been extensively study in the compressed sensing community (generally assuming that $p_{\boldsymbol{x}}$ is a k-sparse model.

Signal Recovery

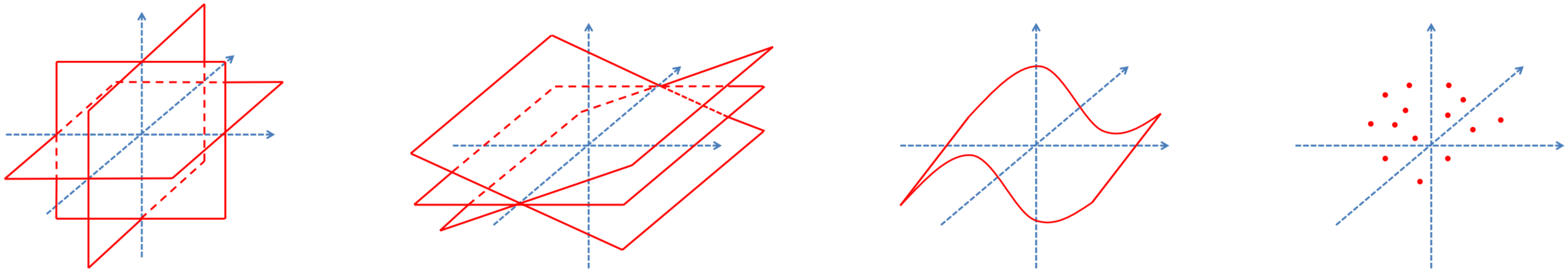
Signal recovery only possible if $\operatorname{supp} p_x = \mathcal{X}$ is **low-dimensional**.

There are multiple ways to 'measure' low-dimensionality.

A popular choice is **box-counting dimension**:

$$\dim(\mathcal{X}) = \lim_{\epsilon \to 0} -\frac{\log N(\mathcal{X}, \epsilon)}{\log \epsilon}$$

where $N(\mathcal{X}, \epsilon)$ is the size of an ϵ-covering of $\mathcal{X}$

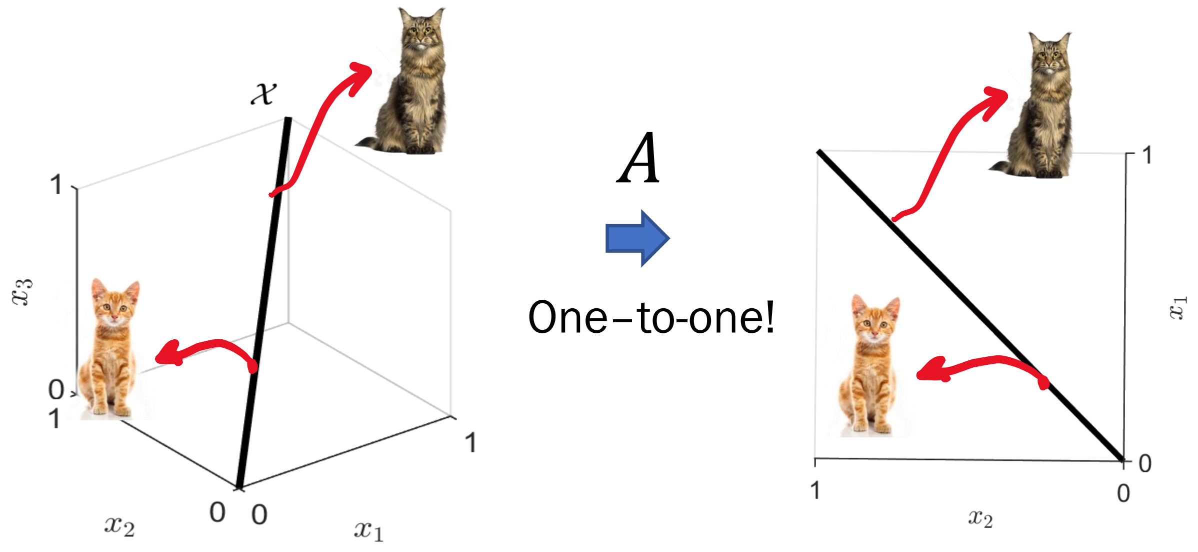


Examples: Sparse dictionaries, manifold models, union-of-subspace models, etc. [Bourrier et al., 2014]

Signal Recovery

Theorem: [Sauer et al., 1991] A signal $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^n$ with $\dim(\mathcal{X}) = k$ can be uniquely recovered from $\boldsymbol{y} = A\boldsymbol{x}$ with almost every $A \in \mathbb{R}^{m \times n}$ if

$$m > 2k.$$



Model Identification

- Model identification is a **linear** inverse problem in **infinite** dimensions

$$p_y(\boldsymbol{y}) = \int p(\boldsymbol{y}|\boldsymbol{x}) p_{\boldsymbol{x}}(\boldsymbol{x}) d\boldsymbol{x}$$

$$\boxed{p_{\boldsymbol{y}} = \boldsymbol{\mathcal{A}}(p_{\boldsymbol{x}})}$$

- Here we assume access to p_y, however, in practice we only have finite observations $\hat{p}_{\boldsymbol{y}} = \sum_{i=1}^{N} \delta_{\boldsymbol{y}_i}$

Can we learn with noise?

Noisy measurement setting $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\epsilon}$

- For additive noise $p(\boldsymbol{y}|\boldsymbol{x}) = g(\boldsymbol{x} - \boldsymbol{y})$:

$$p_{\boldsymbol{y}} = \mathcal{N}\left(0, I\sigma^2\right) * p_{\boldsymbol{x}}$$

- This is a **deconvolution** problem!

- In Fourier we have, $\phi_{\boldsymbol{y}}(\boldsymbol{\omega}) = \phi_{\boldsymbol{x}}(\boldsymbol{\omega})\, \hat{g}(\boldsymbol{\omega})$ where $\phi_{\boldsymbol{x}}$ and $\phi_{\boldsymbol{y}}$ are the characteristic functions of $p_{\boldsymbol{x}}$ and $p_{\boldsymbol{y}}$, and $\hat{g}$ is the Fourier transform of g.

Can we learn with noise?

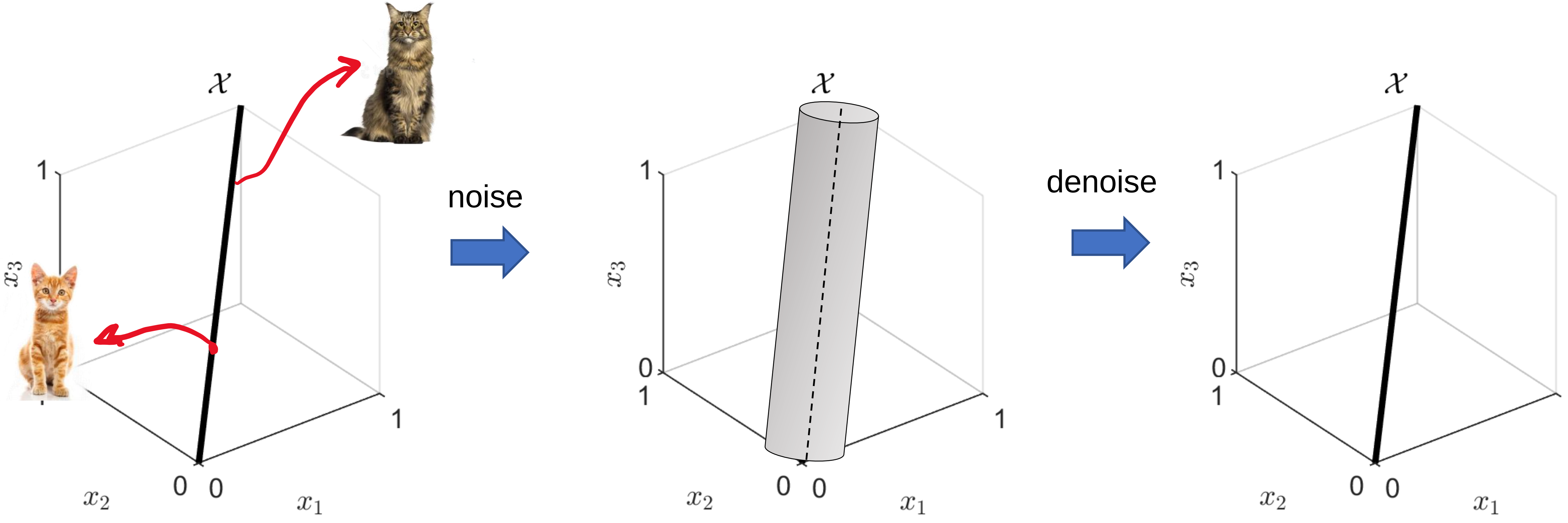
- Since $\mathcal{N}(\mathbf{0}, I\sigma^2)$ is an invertible kernel $\hat{g}(\boldsymbol{\omega}) \neq 0$ for all $\boldsymbol{\omega}$, we can identify $p_{\boldsymbol{x}}$ from $p_{\boldsymbol{y}}$

Proposition [T. et al., 2023]: For additive noise with nowhere zero characteristic function, it is possible to uniquely identify $p_{\boldsymbol{x}}$ from $p_{\boldsymbol{y}}$.

- For non-additive noise (eg. Poisson), the problem is slightly harder

Geometric intuition

Toy example ($n = 3$): Signal set is $\mathcal{X} = \operatorname{supp} p_x = \operatorname{span}[1,1,1]^{\mathrm{T}}$, Gaussian noise.



Incomplete Measurements

We now consider incomplete measurements $\boldsymbol{y}_i = A_{g_i}\boldsymbol{x}_i$ with $g \in \{1, \ldots, G\}$

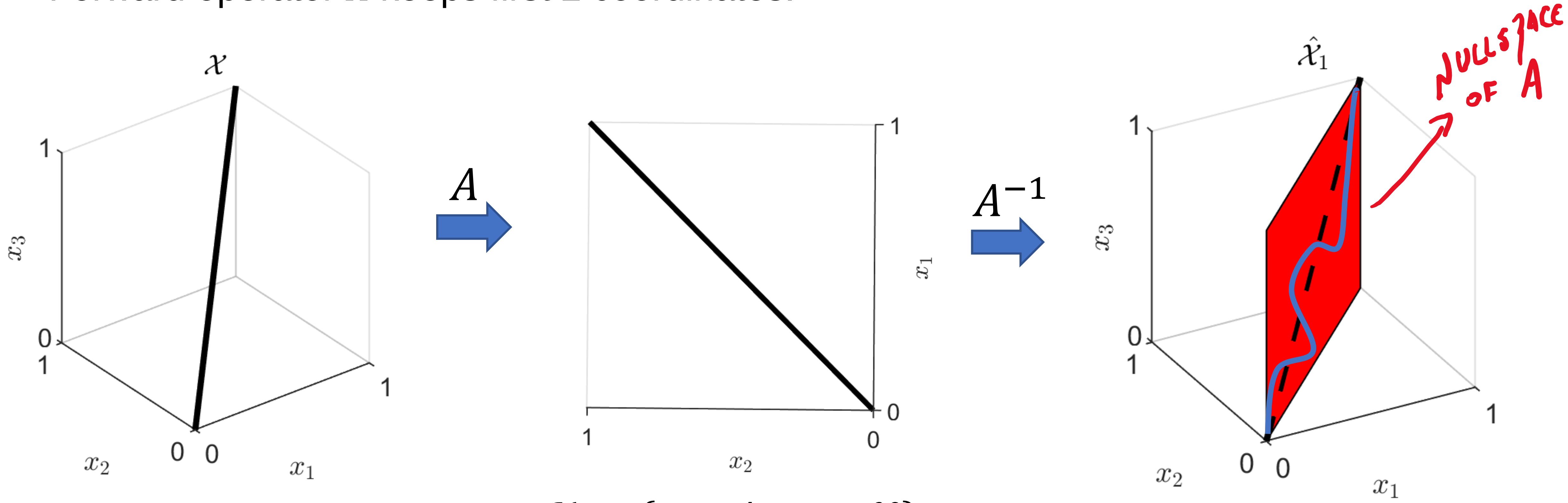
- Either multiple operators, or equivariance $A_g = AT_g$

Can we **uniquely** identify the distribution $p_{\boldsymbol{x}}$ from the measurement distribution $p_{\boldsymbol{y}}$ when the A_g's are incomplete?

- If $p_{\boldsymbol{x}}$ has a low-dimensional support, we can focus on recovering $\operatorname{supp} p_{\boldsymbol{x}} = \mathcal{X}$ from $\operatorname{supp} p_y = \left\{A_g\mathcal{X}\right\}_{g=1:G}$

Geometric intuition

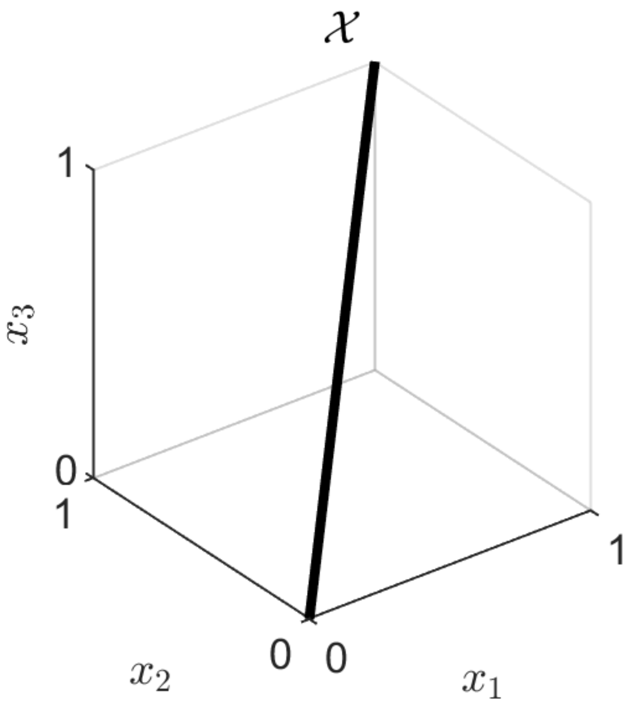
Toy example ($n = 3, m = 2$): Signal set is $\mathcal{X} = \mathrm{span}([1,1,1]^{\mathrm{T}})$
Forward operator A keeps first 2 coordinates.



$$\mathcal{Y} := \{y = Ax, x \in \mathcal{X}\}$$

Geometric intuition

Toy example ($n = 3, m = 2$): Signal set is $\mathcal{X} = \mathrm{span}[1,1,1]^{\mathrm{T}}$. Forward operator A keeps first 2 coordinates. **Now with explicit shift symmetry**



$$\mathcal{X} = \bigcap_{g \in G} A_g^{-1} \mathcal{Y}$$

Sufficient Condition

- Multiple operator setting: assume $A_1, \ldots, A_G$ are **generic**

Theorem [T. et al., 2023]*:* Identifying a k-dimensional $\mathcal{X}$ from observed sets $\{\mathcal{Y}_g = A_g \mathcal{X}\}_{g=1}^{G}$ is possible by almost every $A_1, \ldots A_G \in \mathbb{R}^{m \times n}$ if

$$m > k + \frac{n}{G}$$

- If $G > n$, then the bound is similar to signal recovery.
- ‘almost-every’ result, doesn’t say what happens for a specific subset (eg MRI operators).

Sufficient Condition

- Single operator setting: assume A is **generic**

Theorem [T. et al.]: G cyclic group. Identifying k-dim G-invariant set $\mathcal{X}$ possible by almost every $A \in \mathbb{R}^{m\times n}$ with

$$m > 2k + \max c_j + 1 \geq 2k + \frac{n}{|G|} + 1$$

where c_j is the multiplicity of the representation.

- If $G \;\; > n$, then the bound is similar to signal recovery.
- 'almost-every' result, doesn't say what happens for a specific subset (e.g. MRI operator).

Can we learn any model?

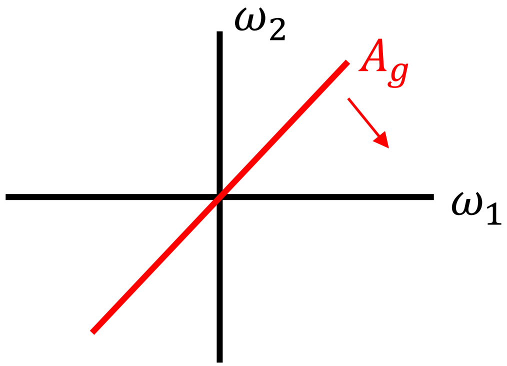
Do we need the low dimensional assumption?

We can analyse identifiability using the characteristic function $\phi_{\boldsymbol{x}}(\boldsymbol{\omega})$

- For finite groups/finite operators, we only observe $\phi_{\boldsymbol{x}}$ in $\cup_{g\in G}\,\mathrm{range}(A_g)$:

$$\mathbb{E}_{\boldsymbol{y}}\, e^{\mathrm{i}\boldsymbol{\omega}^\top A_g^\dagger \boldsymbol{y}} = \mathbb{E}_{\boldsymbol{x}}\, e^{\mathrm{i}\boldsymbol{x}^\top A_g^\dagger A_g \boldsymbol{\omega}} = \phi_{\boldsymbol{x}}(A_g^\dagger A_g \boldsymbol{\omega})$$

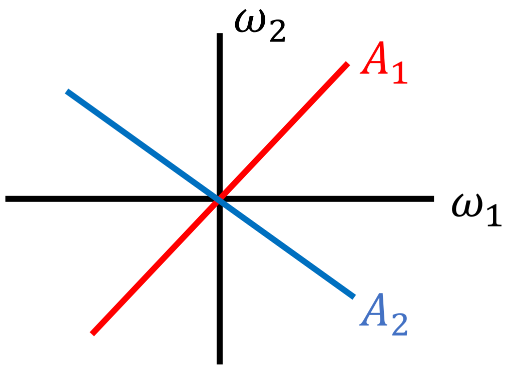
Theorem [Cramer and Wold, 1936]: Any distribution $p_{\boldsymbol{x}}$ *is* uniquely determined by **all** its one-dimensional ($m=1$) projections.



Can we learn any model?

In practice, the Cramer Wold theorem is not verified, as it requires infinitely diverse operators.

- To uniquely identify $p_{\boldsymbol{x}}$ we need that $\bigcup_{g \in G} \mathrm{range}(A_g) = \mathbb{R}^n$ which only holds for $G \to \infty$



References

The full reference list for this tutorial can be found here:

https://tachella.github.io/projects/selfsuptutorial/