



Fundamental performance limits of ideal decoders in high-dimensional linear inverse problems

A. Bourrier^{a,b}, M. Davies^c, T. Peleg^d, P. Perez^b, R. Gribonval^a

a INRIA Rennes - Bretagne Atlantique, France

b Technicolor, Rennes, France

c The University of Edinburgh, UK

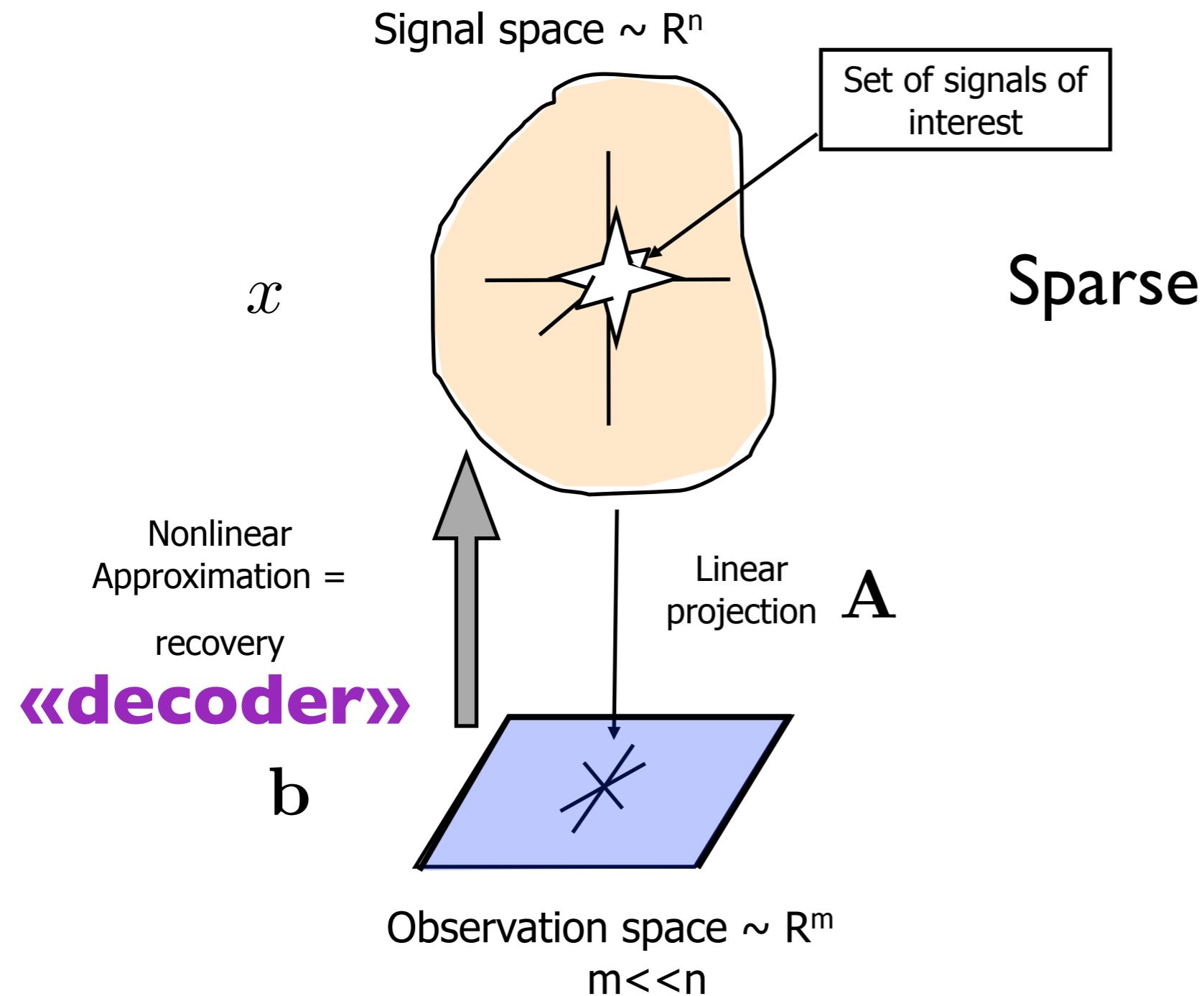
d The Technion, Israel



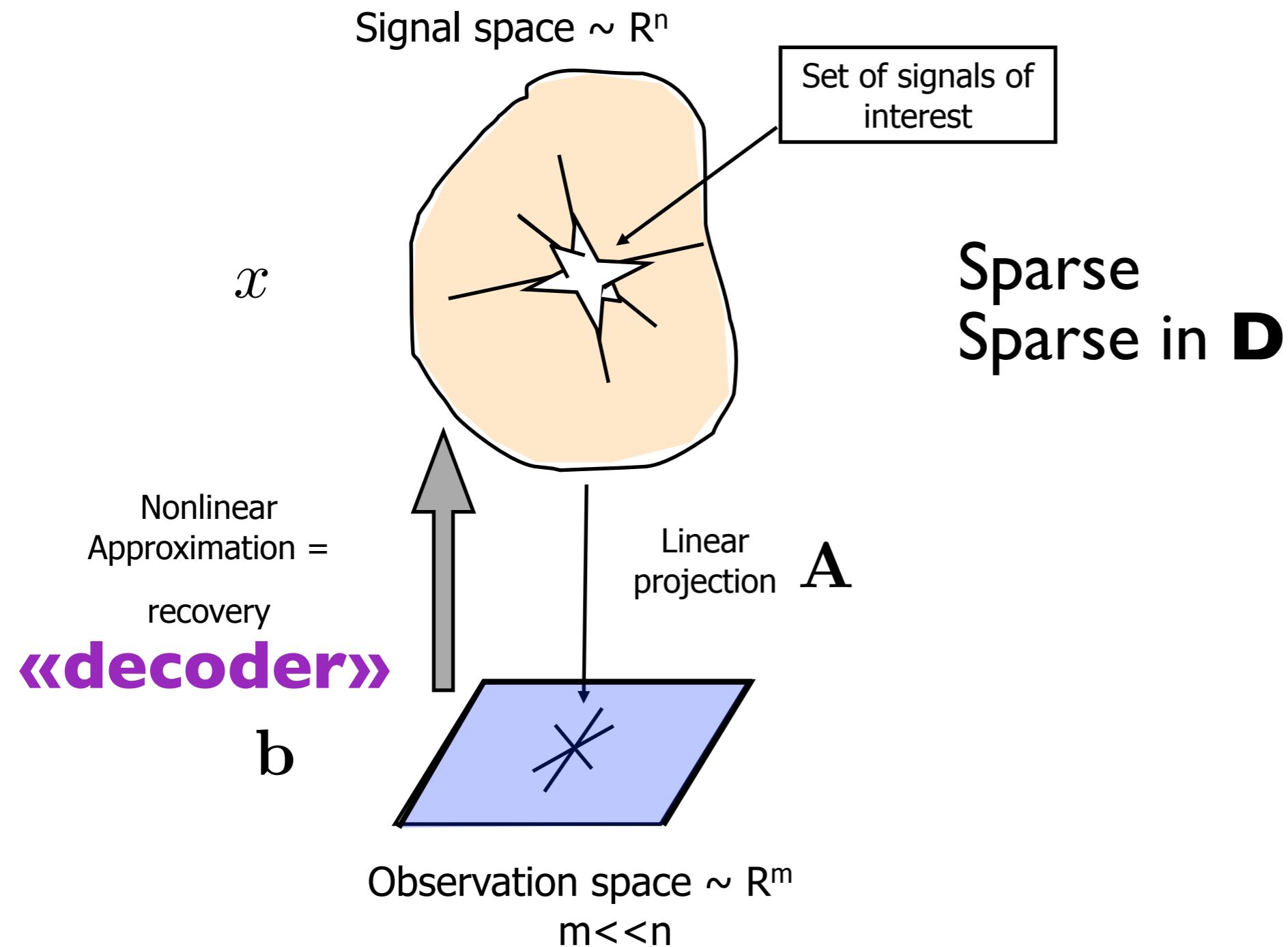
Outline

- Inverse problems & general low-dimensional models
- Noise-robust decoders: noise-blind vs noise-aware
- Dimension reduction ?
- Generalized RIP and Atomic Norms
- Decoders ?

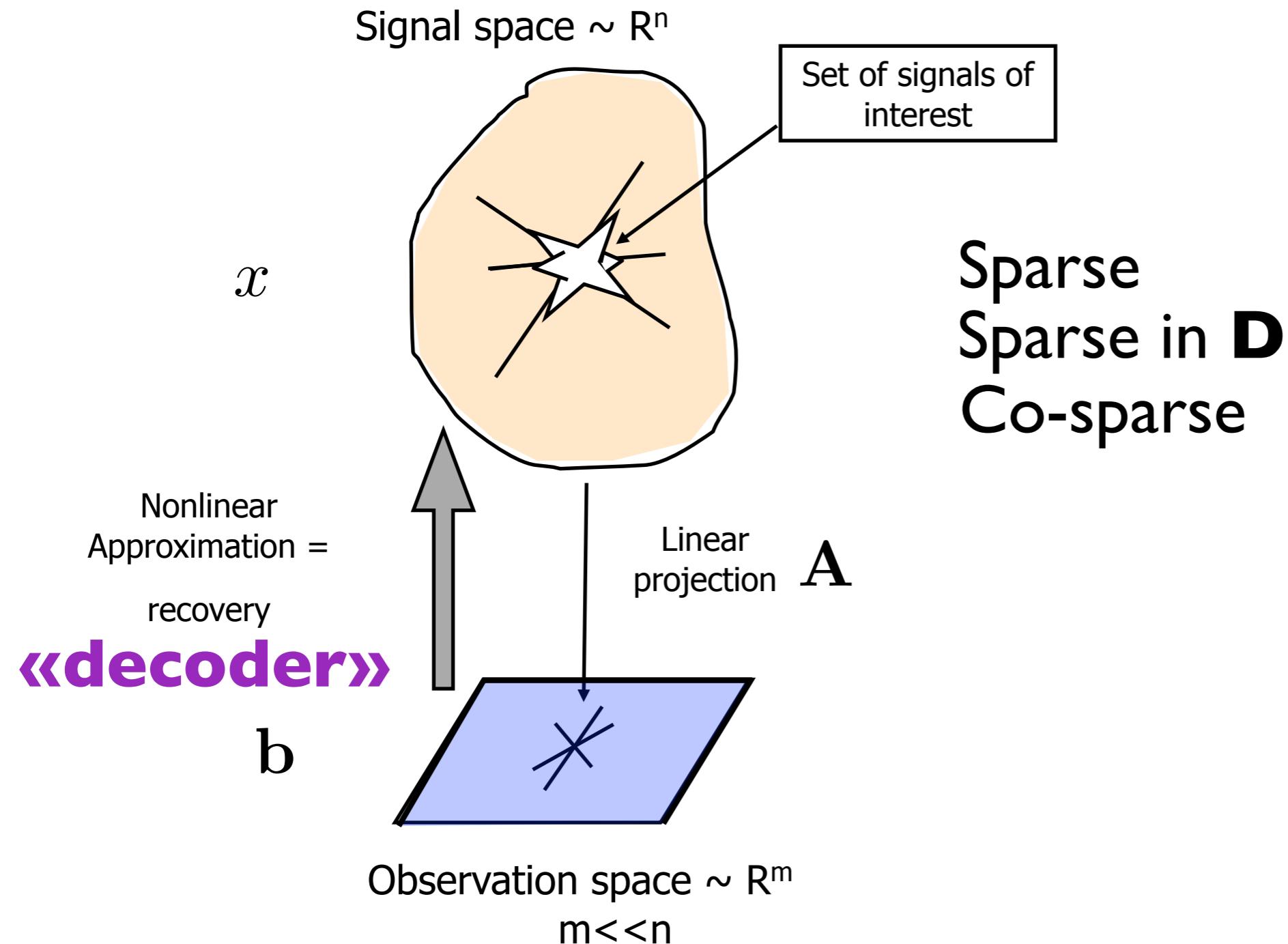
Inverse problems and «sparse» models



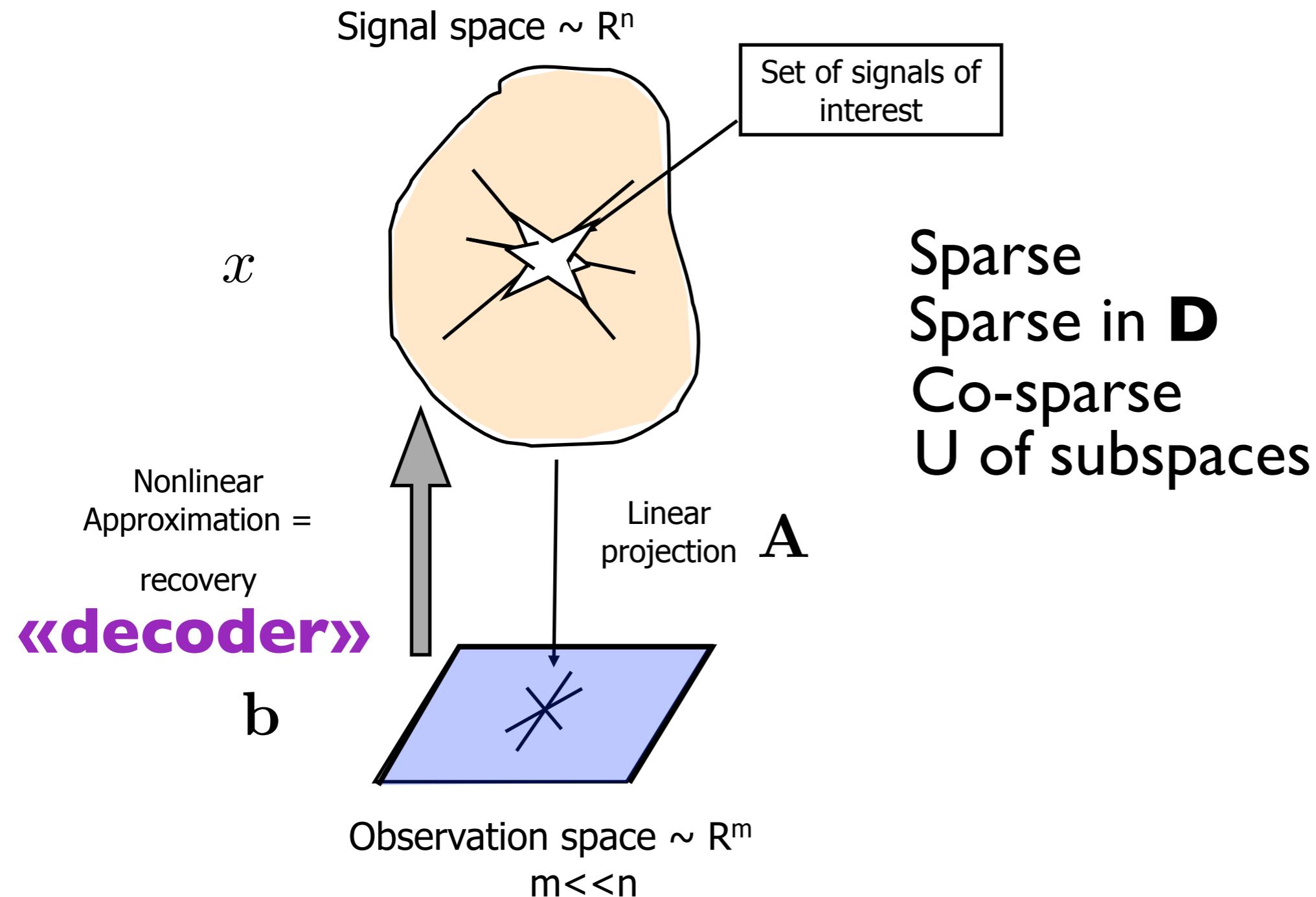
Inverse problems and «sparse» models



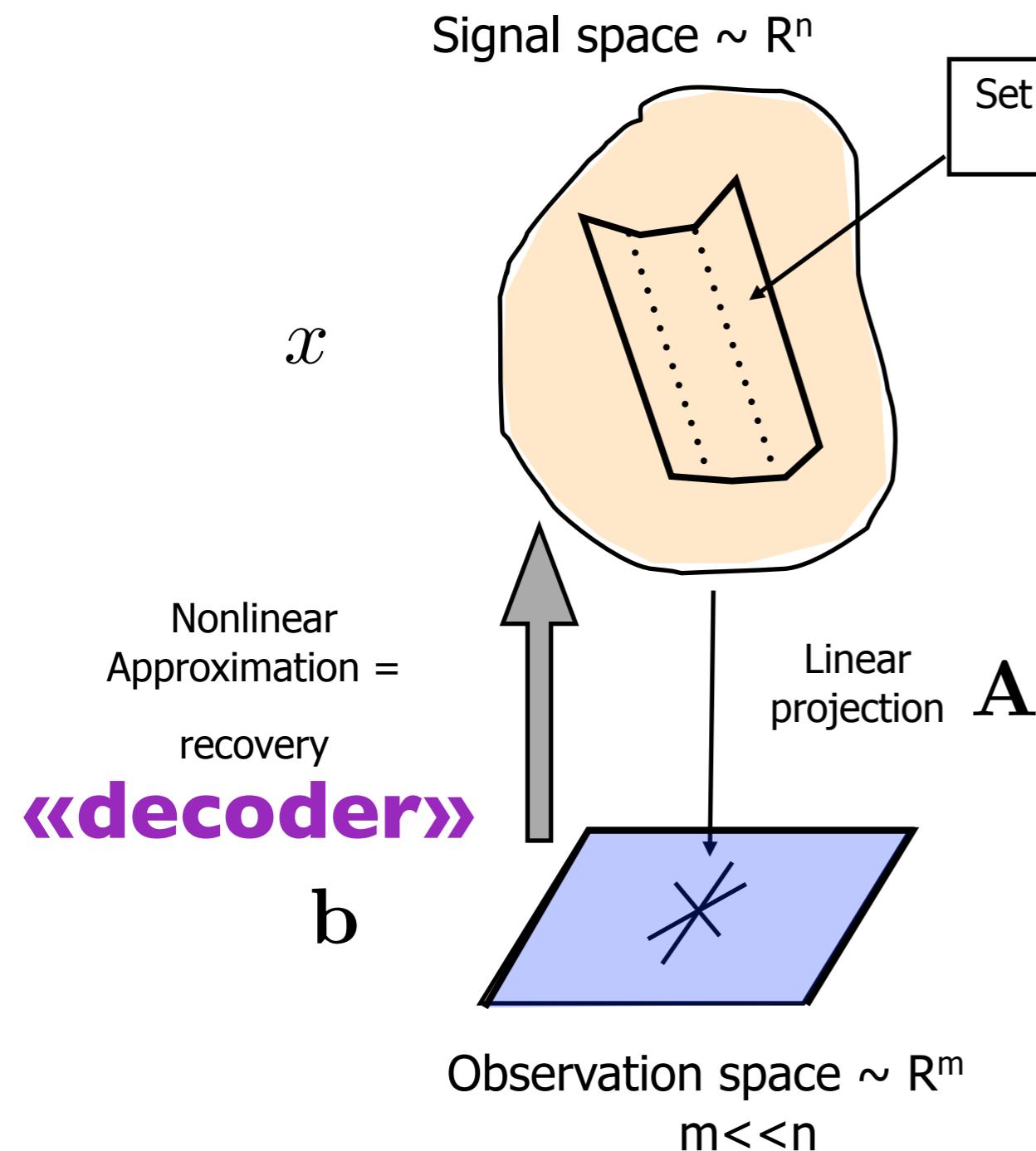
Inverse problems and «sparse» models



Inverse problems and «sparse» models

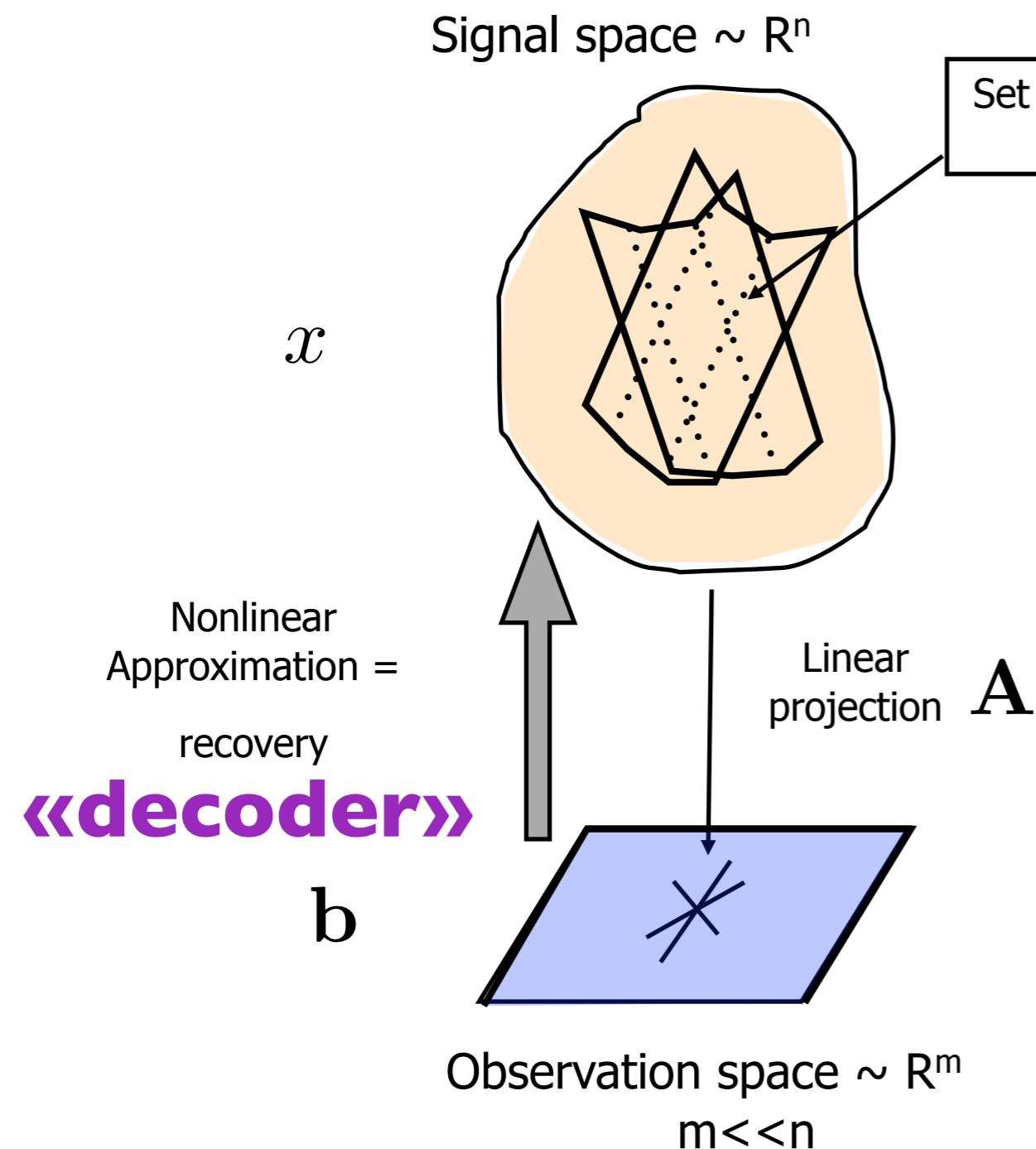


Inverse problems and «sparse» models



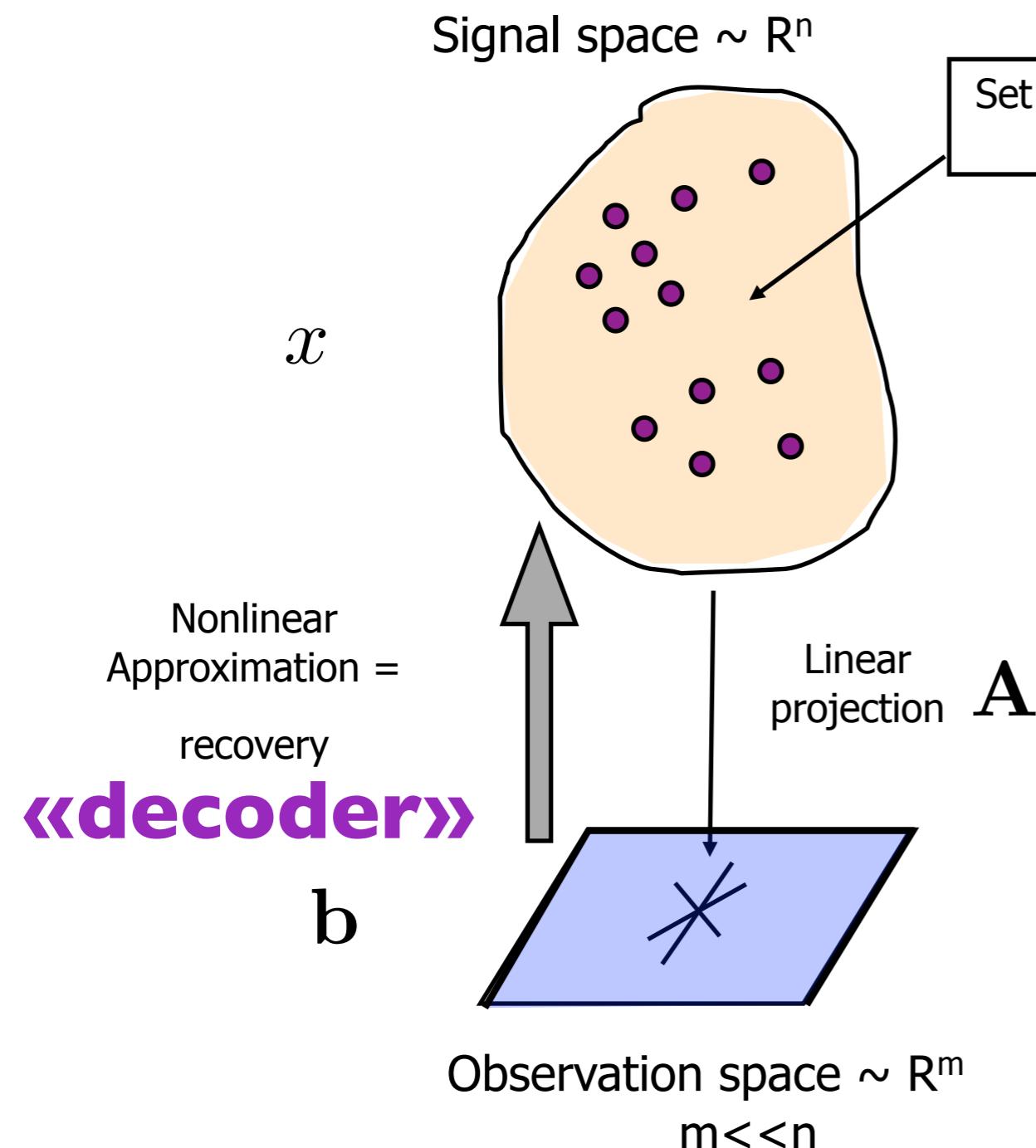
Sparse
Sparse in **D**
Co-sparse
U of subspaces
Low rank matrix
Low rank tensor
Manifold (FRI, off-the grid)

Inverse problems and «sparse» models



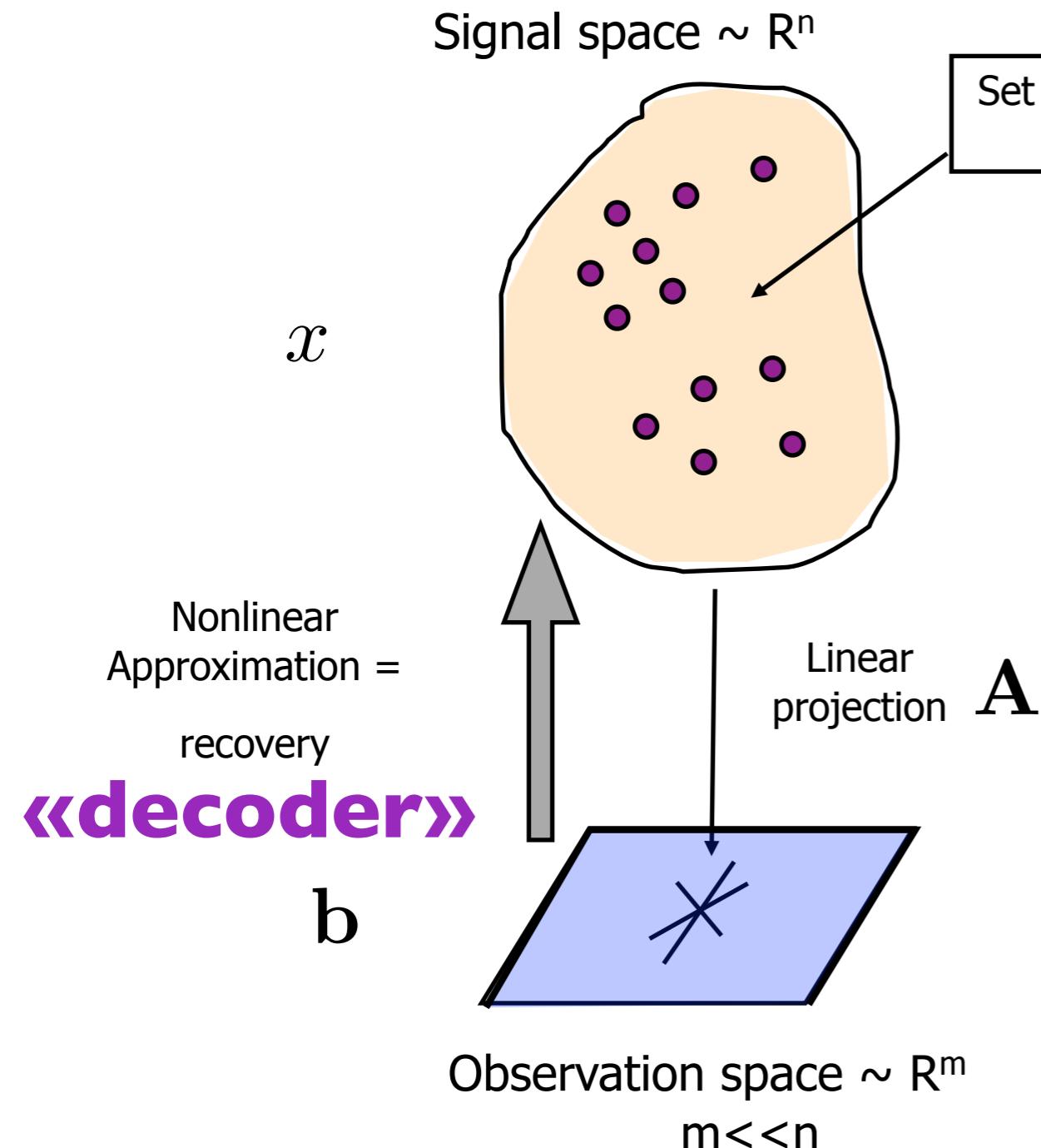
Sparse
Sparse in **D**
Co-sparse
U of subspaces
Low rank matrix
Low rank tensor
Manifold (FRI, off-the grid)
U of manifolds

Inverse problems and «sparse» models



Sparse
Sparse in **D**
Co-sparse
U of subspaces
Low rank matrix
Low rank tensor
Manifold (FRI, off-the grid)
U of manifolds
Point cloud (database)
...

Inverse problems and «sparse» models



$\Sigma \subset \mathbb{R}^n$ **Arbitrary set**

- Sparse
- Sparse in **D**
- Co-sparse
- \cup of subspaces
- Low rank matrix
- Low rank tensor
- Manifold (FRI, off-the grid)
- \cup of manifolds
- Point cloud (database)
- ...

Instance optimal decoders

- **Problem:**

- ✓ given linear measurement operator $\mathbf{A} : E \rightarrow F$
- ✓ given *general model set* $\Sigma \subset E$
- ✓ is there a stable recovery «algorithm» = decoder ?
 $\Delta : F \rightarrow E$

- **Definition:** decoder is *instance optimal* [Cohen & al 2009] with respect to norms $\|\cdot\|_X, \|\cdot\|_Y$ if for any x in E ,

$$\|x - \Delta(\mathbf{A}x)\|_X \leq Cd(x, \Sigma)_Y$$

- **Ex:** with $\Sigma = \Sigma_k$ (*k*-sparse vectors), $d(x, \Sigma) = \sigma_k(x)_Y$

Existence of instance optimal decoder

- **Theorem 1:** equivalence between

- ✓ *Existence of instance optimal decoder*

$$\|x - \Delta(\mathbf{A}x)\|_X \leq Cd(x, \Sigma)_Y$$

- ✓ **Null space property:** for any vector $h \in \text{Ker } \mathbf{A}$

$$\|h\|_X \leq C'd(h, \Sigma - \Sigma)_Y$$

- ◆ with $\Sigma - \Sigma = \{x_1 - x_2, x_i \in \Sigma\}$

- **Ex:** with $\Sigma = \Sigma_k$ the set of k -sparse vectors, $\Sigma - \Sigma = \Sigma_{2k}$

- **What about noise ?**

Outline

- Inverse problems & general low-dimensional models
- **Noise-robust decoders: noise-blind vs noise-aware**
- Dimension reduction ?
- Generalized RIP and Atomic Norms
- Decoders ?

Noise robust instance optimal decoders

- **Question A**

- ✓ Existence of **noise-blind** IO decoder
 - ◆ a single decoder $\Delta : F \rightarrow E$ satisfying

$$\|x - \Delta(\mathbf{A}x + e)\|_X \leq Cd(x, \Sigma)_Y + D\|e\|_Z$$

- ★ for any signal \mathbf{x}
- ★ for any noise \mathbf{e}

- **Question B**

- ✓ Existence of **noise-aware** IO decoder
 - ◆ a family of decoders $\Delta_\epsilon : F \rightarrow E$ such that for any $\epsilon > 0$

$$\|x - \Delta_\epsilon(\mathbf{A}x + e)\|_X \leq Cd(x, \Sigma)_Y + D\epsilon$$

- ★ for any signal \mathbf{x}
- ★ for **small enough** noise $\|e\|_Z \leq \epsilon$

Noise robust instance optimal decoders

- **Question A**

- ✓ Existence of **noise-blind IO decoder**
 - ◆ a single decoder $\Delta : F \rightarrow E$ satisfying

$$\|x - \Delta(\mathbf{A}x + e)\|_X \leq Cd(x, \Sigma)_Y + D\|e\|_Z$$

- ★ for any signal \mathbf{x}
- ★ for any noise \mathbf{e}

- **Theorem 3: equivalence between**

- ✓ Existence of **noise-aware IO decoder**
 - ◆ a family of decoders $\Delta_\epsilon : F \rightarrow E$ such that for any $\epsilon > 0$

$$\|x - \Delta_\epsilon(\mathbf{A}x + e)\|_X \leq Cd(x, \Sigma)_Y + D\epsilon$$

- ★ for any signal \mathbf{x}
- ★ for **small enough** noise $\|e\|_Z \leq \epsilon$



- ✓ «**Robust null space property**»: for any vector h

$$\|h\|_X \leq C'd(h, \Sigma - \Sigma)_Y + D'\|\mathbf{A}h\|_Z$$

reminiscent of [Foucart, Rauhut 2012]

Noise robust instance optimal decoders

- **Theorem 2:** equivalence between

- ✓ Existence of **noise-blind IO decoder**
 - ◆ a single decoder $\Delta : F \rightarrow E$ satisfying

$$\|x - \Delta(\mathbf{A}x + e)\|_X \leq Cd(x, \Sigma)_Y + D\|e\|_Z$$

- ★ for any signal \mathbf{x}
- ★ for any noise \mathbf{e}

- **Theorem 3:** equivalence between

- ✓ Existence of **noise-aware IO decoder**
 - ◆ a family of decoders $\Delta_\epsilon : F \rightarrow E$ such that for any $\epsilon > 0$

$$\|x - \Delta_\epsilon(\mathbf{A}x + e)\|_X \leq Cd(x, \Sigma)_Y + D\epsilon$$

- ★ for any signal \mathbf{x}
- ★ for **small enough** noise $\|e\|_Z \leq \epsilon$



- ✓ «**Robust null space property**»: for any vector h

$$\|h\|_X \leq C'd(h, \Sigma - \Sigma)_Y + D'\|\mathbf{A}h\|_Z$$



reminiscent of [Foucart, Rauhut 2012]

Noise robust instance optimal decoders

- **Theorem 2:** equivalence between

- ✓ Existence of **noise-blind IO decoder**
 - ◆ a single decoder $\Delta : F \rightarrow E$ satisfying

$$\|x - \Delta(\mathbf{A}x + e)\|_X \leq Cd(x, \Sigma)_Y + D\|e\|_Z$$

- ★ for any signal \mathbf{x}
- ★ for any noise \mathbf{e}

- **Theorem 3:** equivalence between

- ✓ Existence of **noise-aware IO decoder**
 - ◆ a family of decoders $\Delta_\epsilon : F \rightarrow E$ such that for any $\epsilon > 0$

$$\|x - \Delta_\epsilon(\mathbf{A}x + e)\|_X \leq Cd(x, \Sigma)_Y + D\epsilon$$

- ★ for any signal \mathbf{x}
- ★ for **small enough** noise $\|e\|_Z \leq \epsilon$

up to constants

- ✓ «Robust null space property»: for any vector h

$$\|h\|_X \leq C'd(h, \Sigma - \Sigma)_Y + D'\|\mathbf{A}h\|_Z$$

reminiscent of [Foucart, Rauhut 2012]

Outline

- Inverse problems & general low-dimensional models
- Noise-robust decoders: noise-blind vs noise-aware
- Dimension reduction ?
- Generalized RIP and Atomic Norms
- Decoders ?

L2/L2 instance optimality ?

- In applications, L2/L2 would be convenient

$$\|x - \Delta(\mathbf{A}x)\|_2 \leq Cd(x, \Sigma)_2$$

- Necessary condition: for all $h \in \text{Ker } \mathbf{A}$

$$\|h\|_X \leq C'd(h, \Sigma - \Sigma)_Y$$

- **Bottleneck** : small C **or** dimension reduction $m \ll n$
 - ✓ with k-sparse model Σ_k [Cohen, Dahmen & De Vore 2009]
 - ✓ with general unions of subspaces Σ [Peleg, G., Davies 2013]
- What about **general sets** Σ ?

Limits of L2/L2 instance optimality

- **Property:** consider $\mathbf{A} : \mathbb{R}^n \rightarrow \mathbb{R}^m$

- ✓ Assume: existence of instance optimal decoder

$$\|x - \Delta(\mathbf{A}x)\|_2 \leq Cd(x, \Sigma)_2$$

- ✓ Assume: Σ contains an orthonormal basis

- ✓ Then $m \geq n/C^2$

- **Generalizations :**

- ✓ e.g., when Σ contains tight frame with $\mathcal{O}(n)$ columns and controlled lower frame bound

Examples

- **k -sparse model / pruned sparse / block-sparse ...**
 - ✓ canonical basis
- **Co-sparse model with finite differences operator**
 - ✓ canonical basis
- **Low-rank / Low-rank & sparse / Low-rank + sparse ...**
 - ✓ canonical basis matrix
- **Low-rank & *anti-sparse***
 - ✓ tensor-product of Fourier bases
- **Covariance matrices with sparse inverse (appear in graphical models)**
 - ✓ canonical basis for symmetric matrices

Replacement for L2/L2 ?

- **Bottleneck:** uniform L2/L2 instance optim. = no future ...
- **Way out 1:** instance optimality **in probability**
 - ✓ see e.g. [Cohen & al 2009], [Chandrasekaran & al 2012]
- **Way out 2:** $L2/\|\cdot\|_Y$ instance optimality
 - ✓ ex: with k -sparse model, under RIP $\|\cdot\|_Y = \|\cdot\|_1/\sqrt{k}$
 - ✓ ex: with rank- r model, under RIP $\|\cdot\|_Y = \|\cdot\|_\star/\sqrt{r}$
- **Extension to more general models?**

Outline

- Inverse problems & general low-dimensional models
- Noise-robust decoders: noise-blind vs noise-aware
- Dimension reduction ?
- **Generalized RIP and Atomic Norms**
- Decoders ?

Model-specific RIP

- **Definition :** $(\Sigma - \Sigma) - RIP$

$$\alpha \|z\|_X \leq \|\mathbf{A}z\|_Z \leq \beta \|z\|_X, \quad z \in \Sigma - \Sigma$$

lower RIP

- **Examples with** $\|\cdot\|_X = \|\cdot\|_Z = \|\cdot\|_2$

- ✓ $\Sigma = \Sigma_k$: standard RIP
- ✓ $\Sigma = \mathbf{D}\Sigma_k$: D-RIP [Candès & al 2011]
- ✓ $\Sigma = \Omega$ – co-sparse : Ω -RIP [Giryes & al 2013]
- ✓ UoS RIP [Blumensath 2011] ...

Model-specific RIP

- **Definition :** $(\Sigma - \Sigma) - RIP$

$$\alpha \|z\|_X \leq \|\mathbf{A}z\|_Z \leq \beta \|z\|_X, \quad z \in \Sigma - \Sigma$$

lower RIP ← **Necessary** for existence of noise-robust decoder
(by Robust Null Space Property)

- **Examples with** $\|\cdot\|_X = \|\cdot\|_Z = \|\cdot\|_2$

- ✓ $\Sigma = \Sigma_k$: standard RIP
- ✓ $\Sigma = D\Sigma_k$: D-RIP [Candès & al 2011]
- ✓ $\Sigma = \Omega$ – co-sparse : Ω -RIP [Giryes & al 2013]
- ✓ UoS RIP [Blumensath 2011] ...

L2 / M-norm

- **Definition:** *M-norm (Measurement matrix dependent)*

- ✓ Assume lower $(\Sigma - \Sigma) - RIP$ $\alpha \|z\|_X \leq \|\mathbf{A}z\|_Z, z \in \Sigma - \Sigma$

- ✓ **Define**

$$\|x\|_M := \|x\|_X + \frac{1}{\alpha} \|\mathbf{A}x\|_Z, \forall x$$

- **Property**

- ✓ Assume lower $(\Sigma - \Sigma) - RIP$

- ✓ Then: existence of *noise-blind instance optimal* decoder

$$\|x - \Delta(\mathbf{A}x + e)\|_X \leq 2d(x, \Sigma)_M + \frac{2}{\alpha} \|e\|_Z$$

L2 / M-norm

- **Definition:** *M-norm (Measurement matrix dependent)*

- ✓ Assume lower $(\Sigma - \Sigma) - RIP \quad \alpha \|z\|_X \leq \|\mathbf{A}z\|_Z, z \in \Sigma - \Sigma$

- ✓ **Define**

$$\|x\|_M := \|x\|_X + \frac{1}{\alpha} \|\mathbf{A}x\|_Z, \forall x$$

- **Property**

- ✓ Assume lower $(\Sigma - \Sigma) - RIP$

- ✓ Then: existence of *noise-blind instance optimal* decoder

$$\|x - \Delta(\mathbf{A}x + e)\|_X \leq 2d(x, \Sigma)_M + \frac{2}{\alpha} \|e\|_Z$$

- **Abstract ? Trivial ? Tautology ? Useless ?**

M-norm & classical guarantees

- **With** $\|\cdot\|_X = \|\cdot\|_Z = \|\cdot\|_2$
 - ✓ **Property 1:** with standard k -sparse model, under full RIP
$$\|x\|_M \leq \left(1 + \frac{\beta}{\alpha}\right) \left(\|x\|_2 + \frac{\|x\|_1}{\sqrt{k}}\right)$$
 - ✓ **Property 2:** with standard rank- r model, under full RIP
$$\|X\|_M \leq \left(1 + \frac{\beta}{\alpha}\right) \left(\|X\|_2 + \frac{\|X\|_*}{\sqrt{r}}\right)$$
- In fact .. relations to *atomic norms* [Chandrasekaran & al 2012]

M -norm and atomic norms

- **Definition:** (generalized) atomic norm

$$\|x\|_{\Sigma} := \inf \sum_k \|x_k\|_X \text{ s.t. } \sum_k x_k = x, \quad x_k \in \mathbb{R}\Sigma$$

- **Property:** if $(\Sigma - \Sigma) - RIP$ then

$$\|x\|_M \leq (1 + \beta/\alpha) \|x\|_{\Sigma}, \quad \forall x$$

- **Example,** with $\|\cdot\|_X = \|\cdot\|_Z = \|\cdot\|_2$

✓ for k -sparse vectors $\|x\|_{\Sigma} \leq \|x\|_2 + \|x\|_1/\sqrt{k}$

✓ for rank- r matrices $\|X\|_{\Sigma} \leq \|X\|_2 + \|X\|_1/\sqrt{r}$

Outline

- Inverse problems & general low-dimensional models
- Noise-robust decoders: noise-blind vs noise-aware
- Dimension reduction ?
- Generalized RIP and Atomic Norms
- Decoders ?

Ideal decoder

- Assume the Robust Null Space Property

$$\|h\|_X \leq C'd(h, \Sigma - \Sigma)_Y + D'\|\mathbf{A}h\|_Z$$

- «Ideal» noise-blind decoder $\mathbf{b} = \mathbf{A}x + e$

$$\Delta(\mathbf{b}) := \arg \min_{u \in E} \{C'd(u, \Sigma)_Y + D'\|\mathbf{A}u - \mathbf{b}\|_Z\}$$

- ◆ Adaptations when minimum may not be achieved (e.g., infinite dimension)

- More concrete decoder ?

Ideal decoder (bis)

- **Definition:** projection onto the model Σ (\sim denoising)

$$P_\Sigma(y) \in \arg \min_{\tilde{x} \in \Sigma} \|y - \tilde{x}\|_2$$

- **Definition:** Projected Landweber Algorithm [Blumensath 2011]

✓ Gradient descent $x^{t+1/2} = x^t + \mu \mathbf{A}^T (\mathbf{A}x^t - \mathbf{b})$

✓ Projection $x^{t+1} = P_\Sigma(x^{t+1/2})$

- **Theorem** [Blumensath 2011]:

✓ Instance optimal with M -norm assuming RIP on $\Sigma - \Sigma$ with constants $\beta \leq 1/\mu \leq 1.5\alpha$

Practical decoder ?

- Need to compute projection onto the model Σ

$$P_\Sigma(y) \in \arg \min_{\tilde{x} \in \Sigma} \|y - \tilde{x}\|_2$$

- Examples

- ✓ k-sparse vectors: *hard thresholding* = easy
- ✓ rank-r matrices: *hard SVD thresholding* ~ easy
- ✓ co-sparse vectors:
 - ◆ NP-hard in general [Tillmann, G., Pfetsch 2013]
 - ◆ *dynamic programming* for piecewise constant signals (1D finite difference operator), or piecewise constant & sparse (1D fused LASSO) [Giryes & al 2013]

Conclusions

- **Summary: feasibility of uniform guarantees**

- ◆ Notion of instance optimality: arbitrary model, arbitrary linear operator
- ◆ Equivalent to (robust) null space property
- ◆ If feasible with noise-aware decoder, then ***noise-blind decoder*** exists
- ◆ L2/L2 incompatible with dim. reduction for many models
- ◆ General RIP implies connections with atomic norms
- ◆ Easy decoding ... when projection on model is easy

- **Preprint arXiv:1311.6239**

- **In the pipe...**

- ◆ extensions to shape of **analysis sparsity guarantees** (with G. Kutyniok)
- ◆ application to Compressive GMM estimation (with A. Bourrier)

- **Challenges**

- ◆ Convex decoders with associated atomic norms ?
- ◆ General theory for ***non-uniform*** guarantees ?

