



Parcimonie et traitement du signal et des images

Partie 4 - Au-delà de la parcimonie - Session 3 - WRAP-UP

Nancy Bertin, Cédric Herzet, Aline Roumy, Rémi Gribonval

remi.gribonval@inria.fr

<http://people.irisa.fr/Remi.Gribonval>

From sparse vectors to low-rank matrices

■ Inverse problems and sparsity

- Definitions and measures of sparsity
- The L_p “norms”
- Sparse recovery
- Sparse approximation

■ Well-posedness of the sparse recovery problem: a key result

- intuition: #measurements & sparsity
- result and demonstration

■ Sparse recovery algorithms

- principles
- first algorithms

■ Compressive sensing

■ Inverse problems and low-rank

- Definitions and measures of rank
- The Shatten (spectral) norms
- Low-rank recovery
- Low-rank approximation

■ Well-posedness of the low-rank recovery problem: a key result

- intuition: #measurements & rank
- result and demonstration (*exercise*)

■ Low-rank recovery algorithms

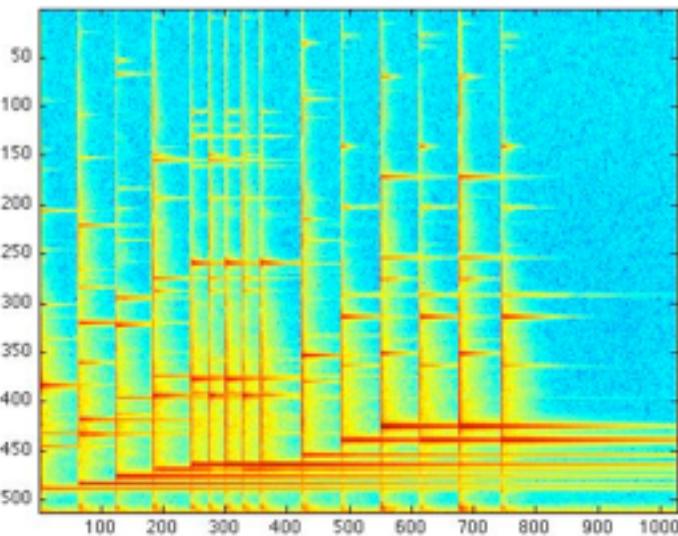
- principles
- first algorithms

■ Compressive Matrix Sensing

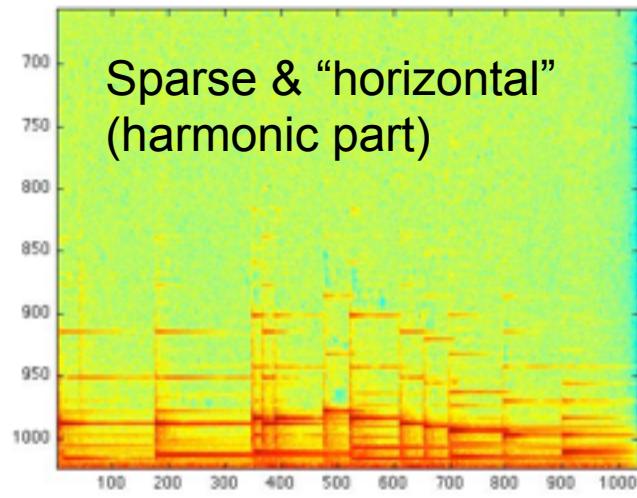
Other low-dimensional models ?
Generic approach / results ?

Example: Structured sparsity

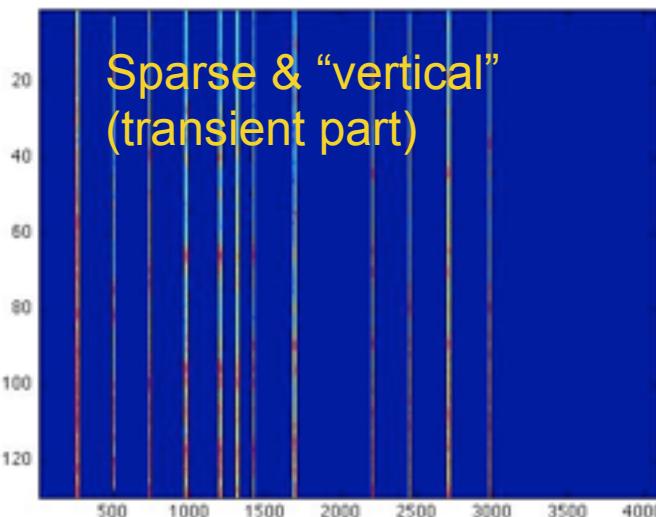
■ Ex: Time-frequency
audio representation



=



+



From sparsity to structured sparsity

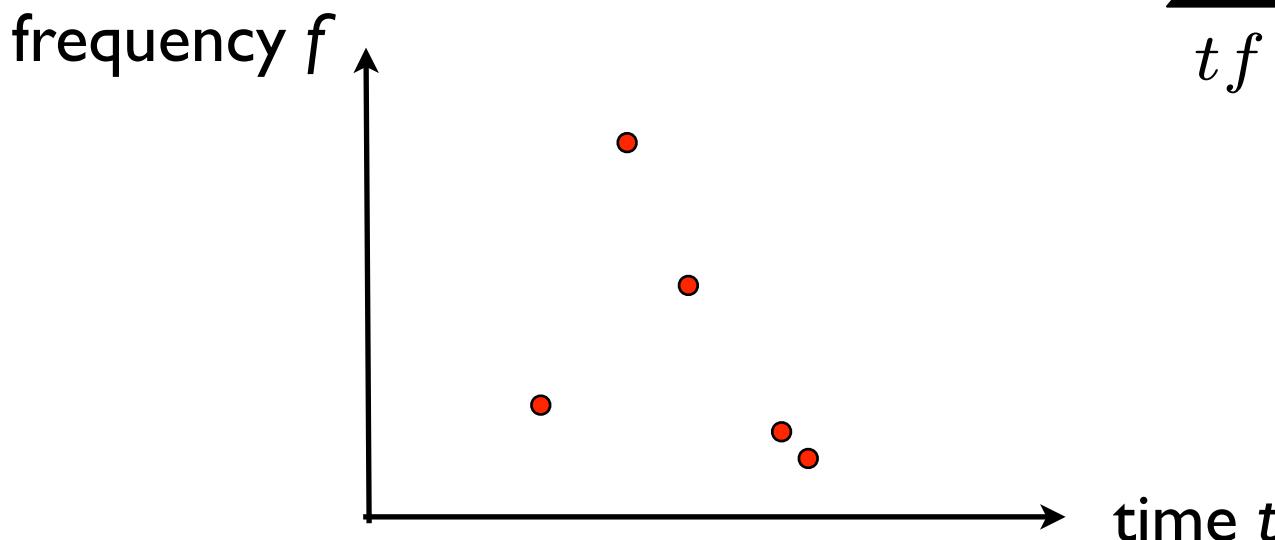
■ Classical sparsity

- measured by

$$\|\mathbf{X}\|_0$$

- convex relaxation:

$$\|\mathbf{X}\|_1 = \sum_{tf} |X_{tf}|$$

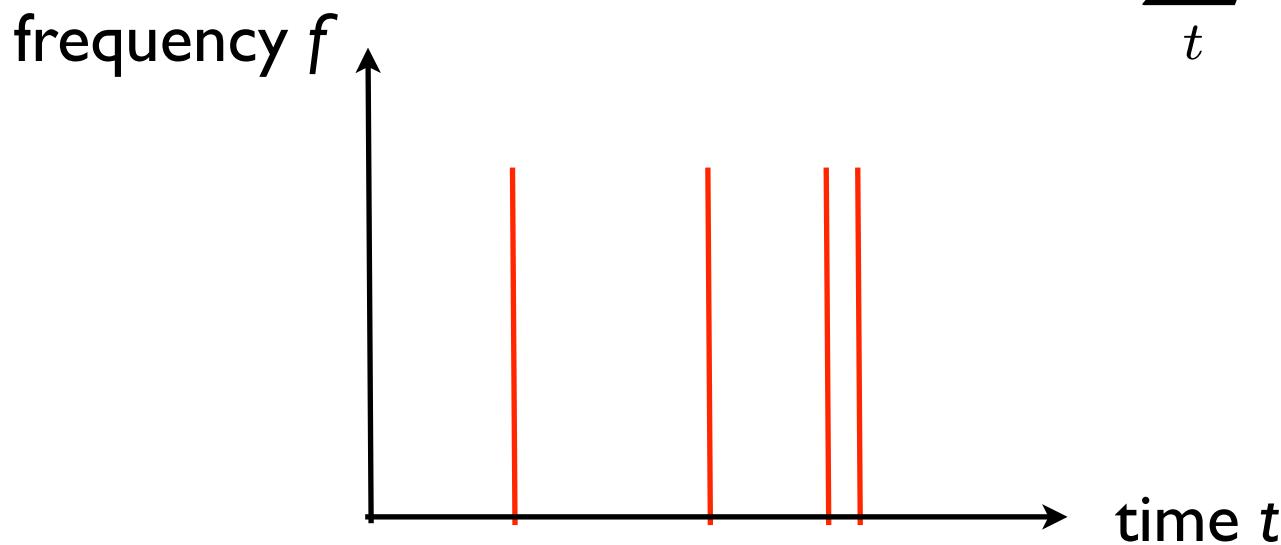


From sparsity to structured sparsity

■ Structured sparsity

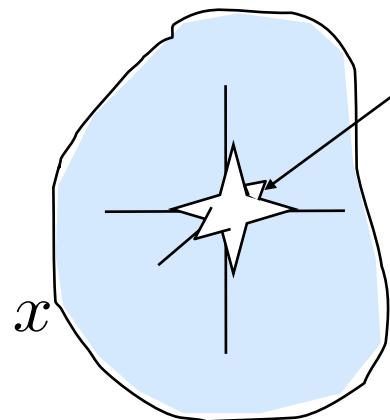
■ measured by $\|\mathbf{X}\|_{\text{col},0} = \#\{t, \|\mathbf{X}(:, t)\|_2 \neq 0\}$

■ convex relaxation: $\|\mathbf{X}\|_{2,1} = \sum_t \|\mathbf{X}(:, t)\|_2$



Stable sparse recovery

Signal space \mathbb{R}^n



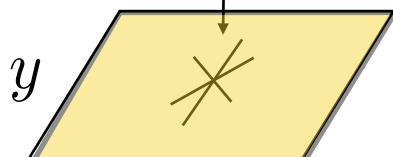
Model set Σ

= signals of interest

set of k-sparse vectors

$$\Sigma_k = \{x \in \mathbb{R}^n, \|x\|_0 \leq k\}$$

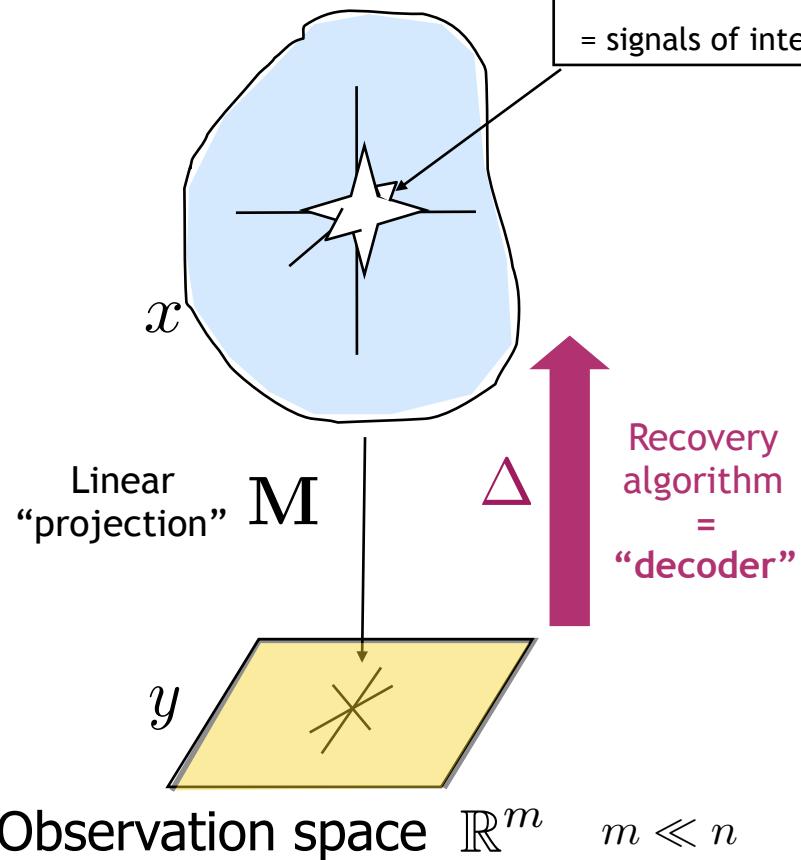
Linear
“projection” \mathbf{M}



Observation space \mathbb{R}^m $m \ll n$

Stable sparse recovery

Signal space \mathbb{R}^n



Model set Σ
= signals of interest

set of k-sparse vectors
 $\Sigma_k = \{x \in \mathbb{R}^n, \|x\|_0 \leq k\}$

Ideal goal: build **decoder** Δ
with the guarantee that

$$\|x - \Delta(\mathbf{M}x + e)\| \leq C\|e\|, \forall x \in \Sigma \quad \forall e$$

(*instance optimality [Cohen & al 2009]*)

Stable recovery of k-sparse vectors

■ “Decoders” studied in this course

- L1 minimization (+optimization algorithms)
■ LASSO [Tibshirani 1994], Basis Pursuit [Chen & al 1999] $\Delta(y) := \arg \min_{x: \mathbf{M}x=y} \|x\|_1$
- Greedy algorithms
 - (Orthonormal) Matching Pursuit [Mallat & Zhang 1993],
 - Iterative Hard Thresholding (IHT) [Blumensath & Davies 2009],
 - ...

■ Guarantees

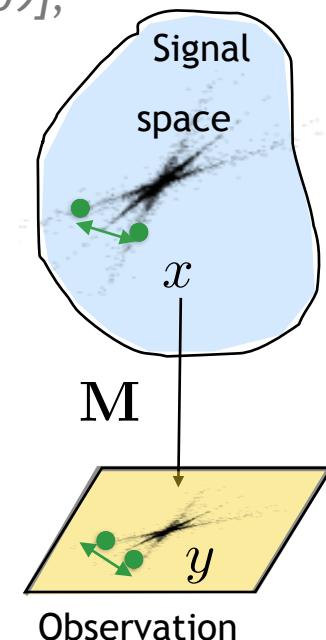
- Assume:
 - \mathbf{M} satisfies the Restricted isometry property (RIP)

[Candès & al 2004]

- Then:
 - Exact recovery
 - Stability to noise
 - Robustness to model error

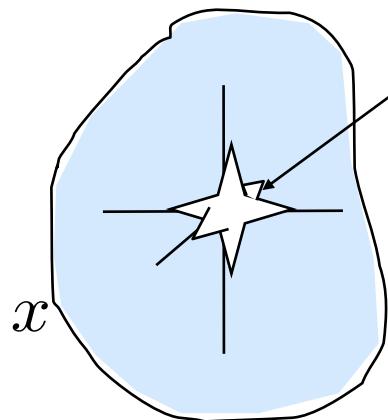
$$1 - \delta \leq \frac{\|\mathbf{M}z\|_2^2}{\|z\|_2^2} \leq 1 + \delta$$

when $\|z\|_0 \leq 2k$



Stable low-rank recovery

Signal space \mathbb{R}^n

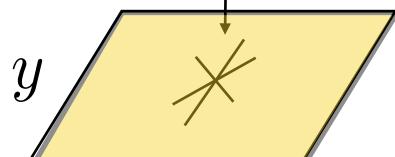


Model set Σ
= signals of interest

set of matrices of rank at most r

$$\Sigma_r = \{\mathbf{X} \in \mathbb{R}^{n \times n}, \text{rank}(\mathbf{X}) \leq r\}$$

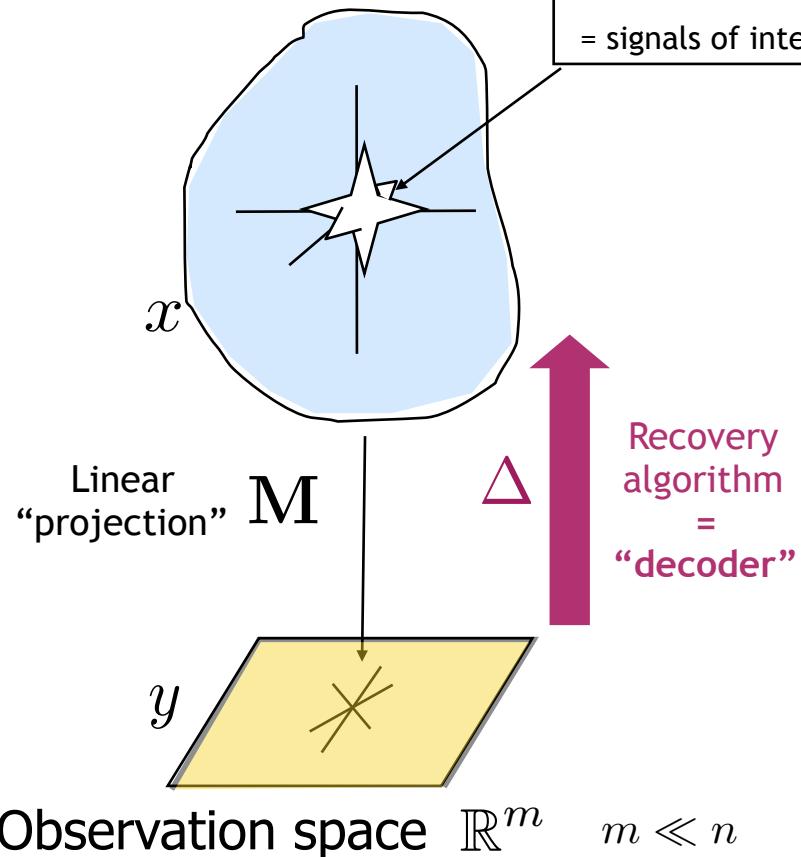
Linear
“projection” \mathbf{M}



Observation space \mathbb{R}^m $m \ll n$

Stable low-rank recovery

Signal space \mathbb{R}^n



Model set Σ
= signals of interest

set of matrices of rank at most r
 $\Sigma_r = \{\mathbf{X} \in \mathbb{R}^{n \times n}, \text{rank}(\mathbf{X}) \leq r\}$

Ideal goal: build **decoder** Δ
with the guarantee that

$\|x - \Delta(\mathbf{M}x + e)\| \leq C\|e\|, \forall x \in \Sigma \quad \forall e$
(instance optimality [Cohen & al 2009])

Stable recovery of low-rank matrices

■ “Decoders” studied in this course

- Nuclear norm minimization (+optimization algorithms)
- Greedy algorithms
 - ADMIRA
 - Singular Value Projection (SVP)
 - ...

$$\Delta(y) := \arg \min_{x: \mathcal{M}(x)=y} \|x\|_*$$

[Lee & Bresler 2009]
[Meka & al 2010],

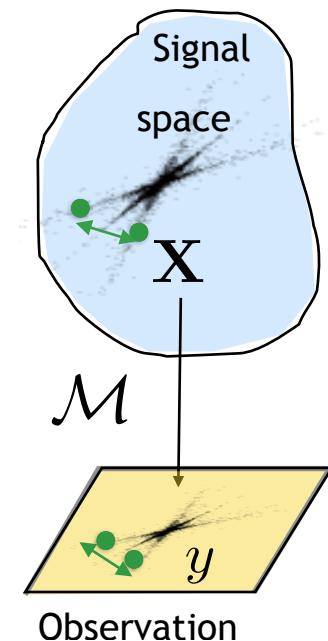
■ Guarantees

- Assume:
 - M satisfies the Restricted isometry property (RIP)

- Then:
 - Exact recovery
 - Stability to noise
 - Robustness to model error

$$1 - \delta \leq \frac{\|\mathcal{M}(Z)\|_2^2}{\|Z\|_F^2} \leq 1 + \delta$$

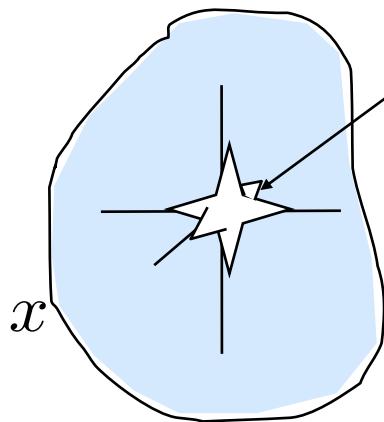
when $\text{rank}(Z) \leq 2r$



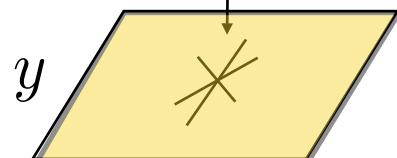
Stable recovery: beyond sparsity or low-rank

Signal space \mathbb{R}^n

Model set Σ
= signals of interest



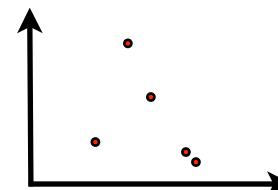
Linear
“projection” \mathbf{M}



Observation space \mathbb{R}^m $m \ll n$

- Low-dimensional model
- Sparse

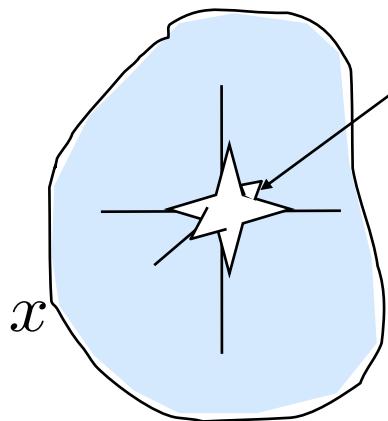
$$\|\mathbf{x}\|_0 \ll n$$



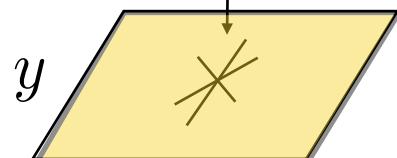
Stable recovery: beyond sparsity or low-rank

Signal space \mathbb{R}^n

Model set Σ
= signals of interest



Linear
“projection” \mathbf{M}

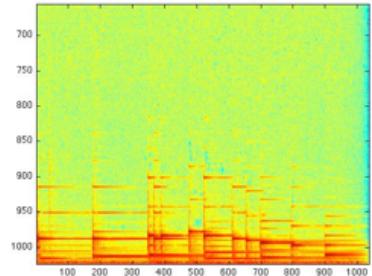
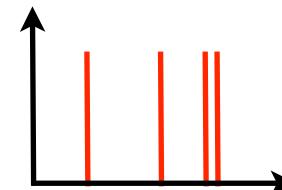


Observation space $\mathbb{R}^m \quad m \ll n$

■ Low-dimensional model

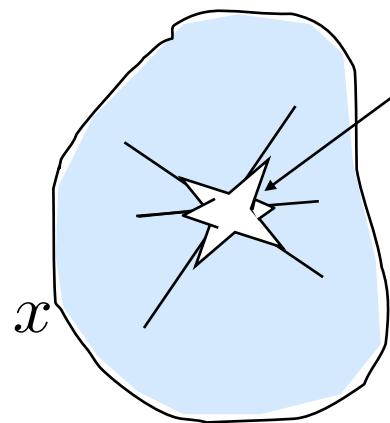
- Sparse
- Structured

$$\|\mathbf{X}\|_{\text{col},0} \ll \sqrt{n}$$



Stable recovery: beyond sparsity or low-rank

Signal space \mathbb{R}^n



Model set Σ

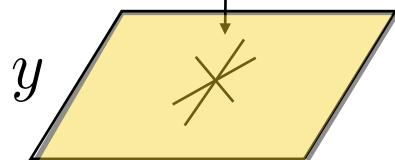
= signals of interest

■ Low-dimensional model

- Sparse
- Structured
- Sparse in dictionary D

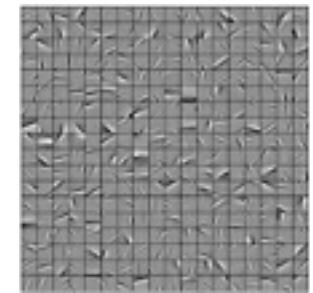
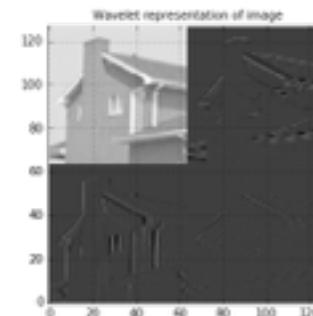
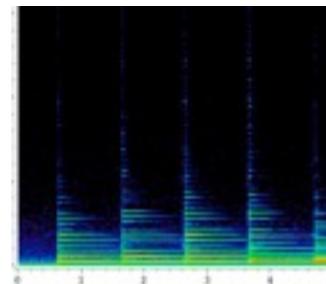
$$\mathbf{x} = \mathbf{Dz}, \quad \|\mathbf{z}\|_0 \ll n$$

Linear
“projection” \mathbf{M}



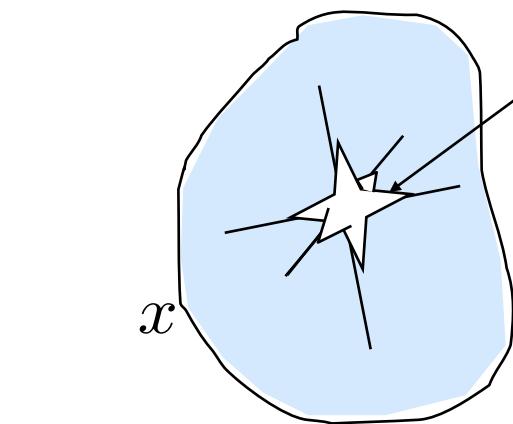
Observation space $\mathbb{R}^m \quad m \ll n$

Dictionary: fixed (Gabor, wavelets ...) or learned

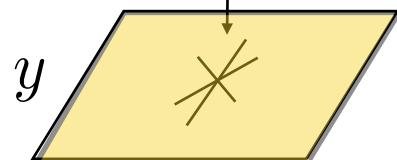


Stable recovery: beyond sparsity or low-rank

Signal space \mathbb{R}^n



Linear
"projection"
 M



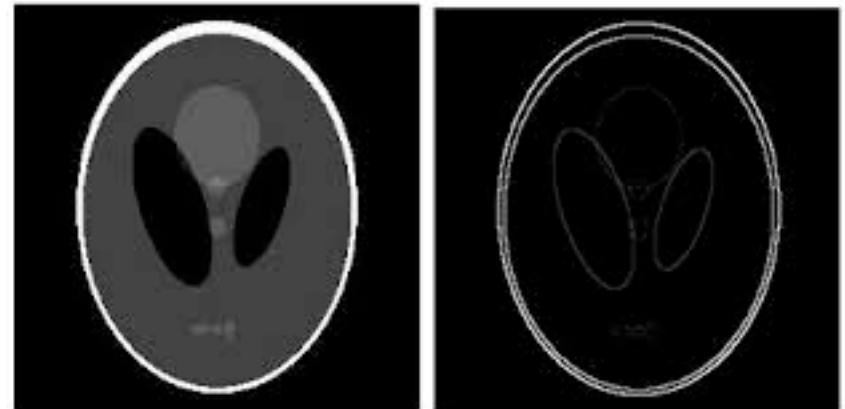
Observation space $\mathbb{R}^m \quad m \ll n$

■ Low-dimensional model

- Sparse
- Structured
- Sparse in dictionary D
- Co-sparse in analysis operator A

$$\|Ax\|_0 \ll n$$

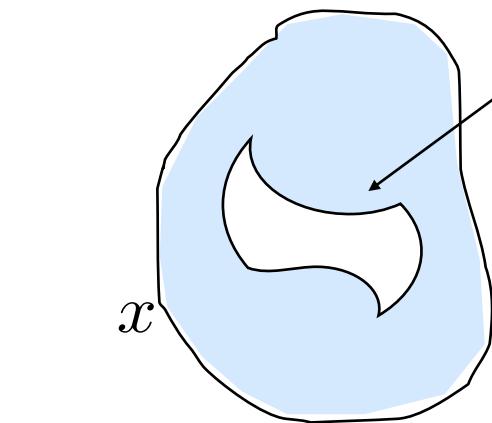
- Ex: small total variation = sparse gradient = piecewise cst



piecewise constant \rightarrow sparse gradient

Stable recovery: beyond sparsity or low-rank

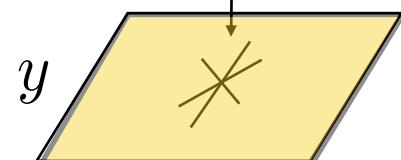
Signal space \mathbb{R}^n



Model set Σ

= signals of interest

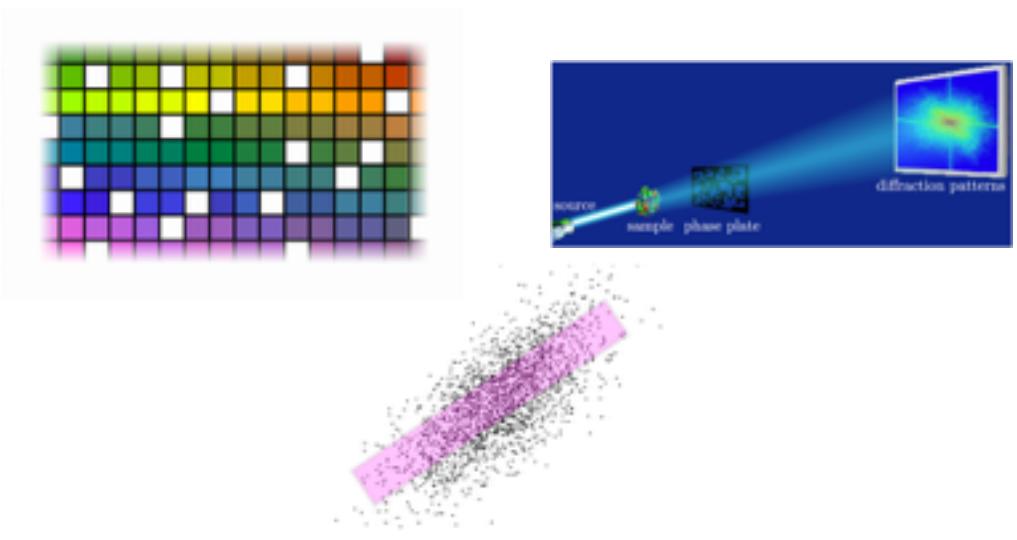
Linear
“projection” \mathbf{M}



Observation space \mathbb{R}^m $m \ll n$

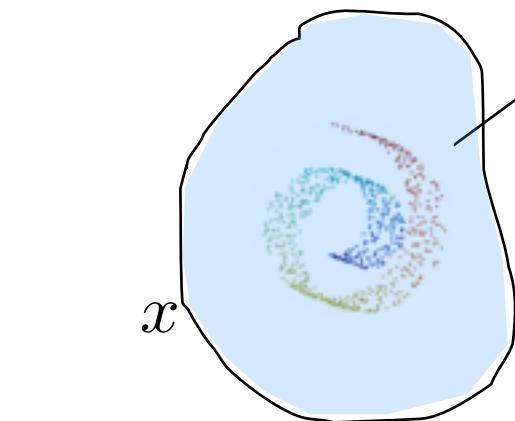
■ Low-dimensional model

- Sparse
- Structured
- Sparse in dictionary \mathbf{D}
- Co-sparse in analysis operator \mathbf{A}
- Low-rank matrix or tensor
 - *matrix completion,*
 - *phase-retrieval,*
 - *compressive PCA*
 - *blind sensor calibration ...*



Stable recovery: beyond sparsity or low-rank

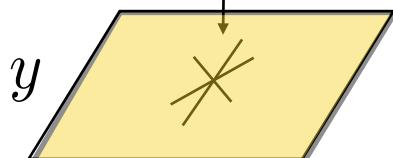
Signal space \mathbb{R}^n



Model set Σ

= signals of interest

Linear
“projection” \mathbf{M}



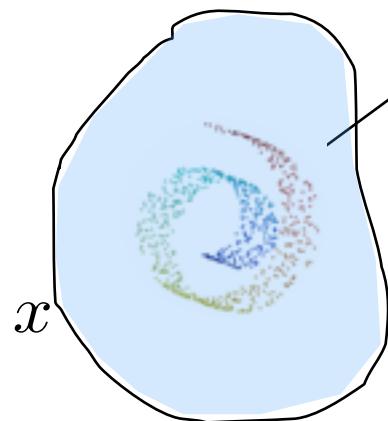
Observation space $\mathbb{R}^m \quad m \ll n$

■ Low-dimensional model

- Sparse
- Structured
- Sparse in dictionary \mathbf{D}
- Co-sparse in analysis operator \mathbf{A}
- Low-rank matrix or tensor
 - *matrix completion,*
 - *phase-retrieval,*
 - *compressive PCA*
 - *blind sensor calibration ...*
- Manifold / Union of manifolds
 - *detection, estimation,*
 - *localization, mapping ...*
- Matrix with sparse inverse
 - *Gaussian graphical models*
- Given point cloud
 - *database indexing*

General stable recovery

Signal space \mathbb{R}^n



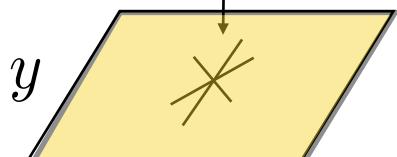
Model set Σ

= signals of interest

■ Low-dimensional model

■ arbitrary set $\Sigma \subset \mathcal{H}$

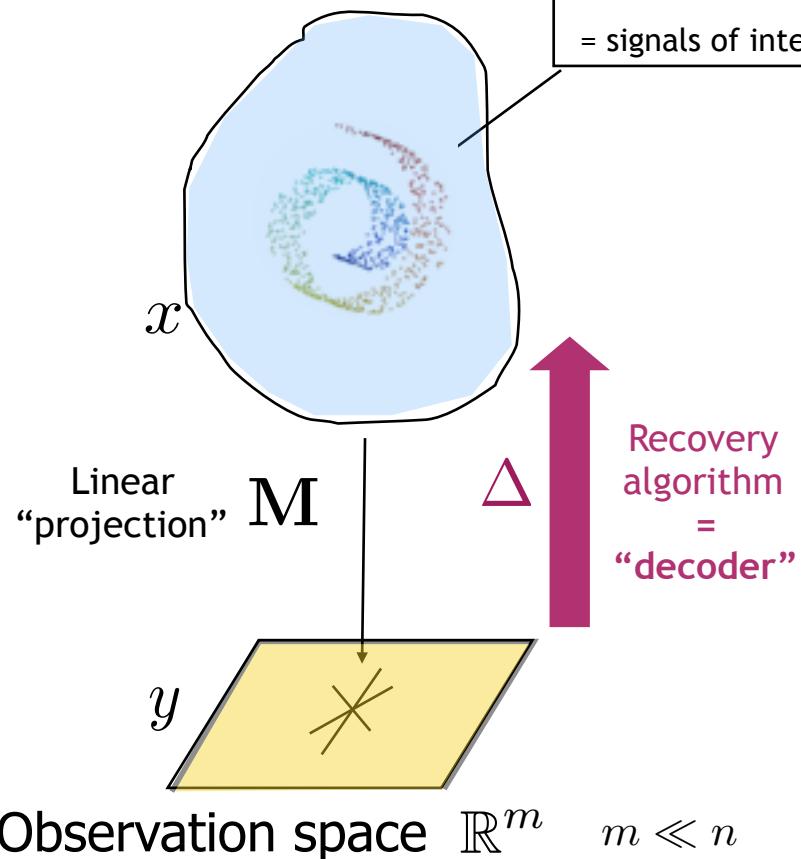
Linear
“projection” \mathbf{M}



Observation space \mathbb{R}^m $m \ll n$

General stable recovery

Signal space \mathbb{R}^n



■ Low-dimensional model

■ arbitrary set $\Sigma \subset \mathcal{H}$

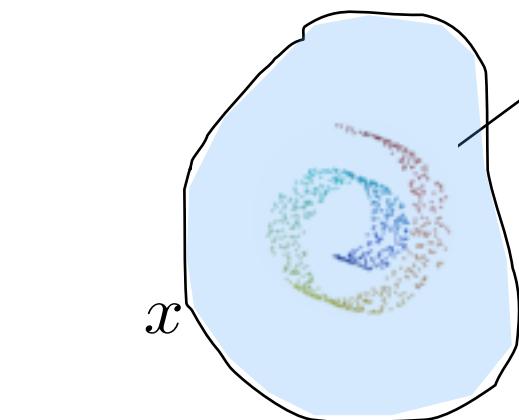
Ideal goal: build decoder Δ with the guarantee that

$$\|x - \Delta(\mathbf{M}x + e)\| \leq C\|e\|, \forall x \in \Sigma \quad \forall e$$

(instance optimality [Cohen & al 2009])

General stable recovery

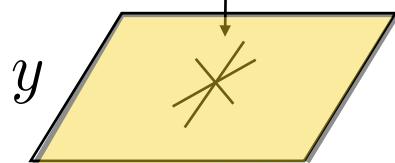
Signal space \mathbb{R}^n



Model set Σ

= signals of interest

Linear
“projection” \mathbf{M}



Observation space \mathbb{R}^m $m \ll n$

■ Low-dimensional model

■ arbitrary set $\Sigma \subset \mathcal{H}$

Ideal goal: build decoder Δ with the guarantee that

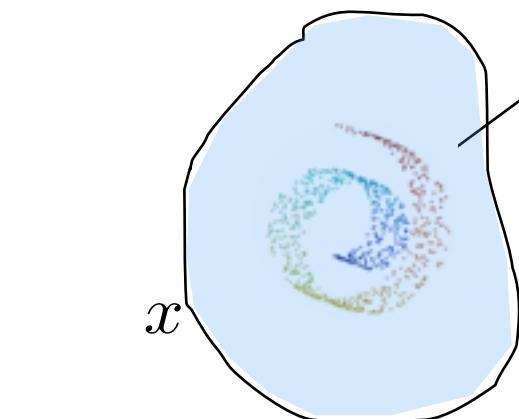
$$\|x - \Delta(\mathbf{M}x + e)\| \leq C\|e\|, \forall x \in \Sigma \quad \forall e$$

(instance optimality [Cohen & al 2009])

Are there such decoders?

General stable recovery

Signal space \mathbb{R}^n



Linear
“projection” \mathbf{M}

Model set Σ

= signals of interest

Recovery
algorithm
= “decoder”

Observation space \mathbb{R}^m $m \ll n$

■ Low-dimensional model

■ arbitrary set $\Sigma \subset \mathcal{H}$

Ideal goal: build decoder Δ with the guarantee that

$$\|x - \Delta(\mathbf{M}x + e)\| \leq C\|e\|, \forall x \in \Sigma \quad \forall e$$

(instance optimality [Cohen & al 2009])

Are there such decoders?

- Literature on embeddings, cf monograph [Robinson 2010]
- [Hurewicz & Wallman '41, Falconer '85, Hunt & Kaloshin '99]

A General Restricted Isometry Property

■ Theorem 1: RIP is necessary

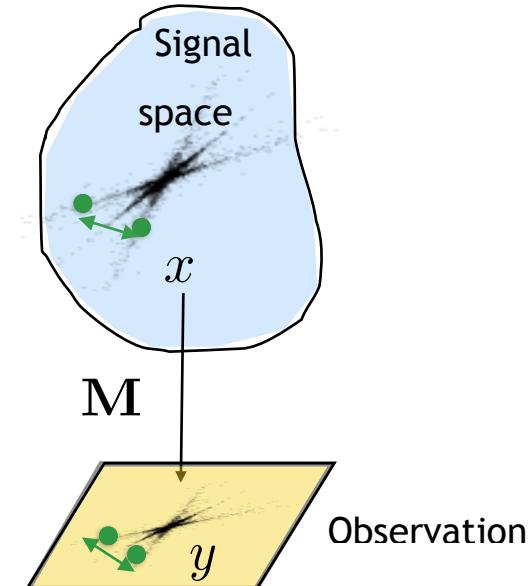
- **Definition:** \mathbf{M} satisfies (general) Restricted Isometry Property (RIP) on **secant set** $\Sigma - \Sigma$ if

$$\alpha \leq \frac{\|\mathbf{M}z\|}{\|z\|} \leq \beta \text{ when } z \in \Sigma - \Sigma := \{x - x', x, x' \in \Sigma\}$$

For example, after renormalization of \mathbf{M}

$$\alpha = \sqrt{1 - \delta}; \beta = \sqrt{1 + \delta}$$

- If *there exists* an instance optimal decoder then \mathbf{M} satisfies the RIP



A General Restricted Isometry Property

■ Theorem 1: RIP is necessary

- **Definition:** \mathbf{M} satisfies (general) Restricted Isometry Property (RIP) on **secant set** $\Sigma - \Sigma$ if

$$\alpha \leq \frac{\|\mathbf{M}z\|}{\|z\|} \leq \beta \text{ when } z \in \Sigma - \Sigma := \{x - x', x, x' \in \Sigma\}$$

For example, after renormalization of \mathbf{M}

$$\alpha = \sqrt{1 - \delta}; \beta = \sqrt{1 + \delta}$$

- If *there exists* an instance optimal decoder then \mathbf{M} satisfies the RIP

■ Theorem 2: RIP is sufficient

- If \mathbf{M} satisfies the RIP then *there exists* an instance optimal decoder:
 - Exact recovery
 - Stable to noise

$$\|x - \Delta(\mathbf{M}x + e)\| \leq C(\delta)\|e\| \quad \forall x \in \Sigma \quad \forall e$$

[Cohen & al 2009] for Σ_k

[Bourrier & al 2014] for arbitrary model set Σ

A General Restricted Isometry Property

■ Theorem 1: RIP is necessary

- **Definition:** \mathbf{M} satisfies (general) Restricted Isometry Property (RIP) on **secant set** $\Sigma - \Sigma$ if

$$\alpha \leq \frac{\|\mathbf{M}z\|}{\|z\|} \leq \beta \text{ when } z \in \Sigma - \Sigma := \{x - x', x, x' \in \Sigma\}$$

For example, after renormalization of \mathbf{M}

$$\alpha = \sqrt{1 - \delta}; \beta = \sqrt{1 + \delta}$$

- If *there exists* an instance optimal decoder then \mathbf{M} satisfies the RIP

■ Theorem 2: RIP is sufficient

- If \mathbf{M} satisfies the RIP then *there exists* an instance optimal decoder:
 - Exact recovery
 - Stable to noise
 - Bonus: robust to model error

$$\|x - \Delta(\mathbf{M}x + e)\| \leq C(\delta)\|e\| + C'(\delta)d_{\Sigma}(x, \Sigma)$$

[Cohen & al 2009] for Σ_k

[Bourrier & al 2014] for arbitrary model set Σ

Distance to model set

Ideal decoder

■ Instance optimality guarantee under RIP

$$\|x - \Delta(\mathbf{M}x + e)\| \leq C(\delta)\|e\| + C'(\delta)d_{\Sigma}(x, \Sigma)$$

■ «Ideal & Universal» instance optimal decoder

$$\Delta(y) := \arg \min_{x \in \Sigma} \|\mathbf{M}x - y\|$$

- NB: minimum may not be achieved (e.g. infinite dimension) - definition can be adapted.
- + Noise-blind: no knowledge of noise level needed
- - Not very practical ...

Greedy decoders

■ Projected Landweber Algorithm (PLA)

- generalizes Iterative Hard Thresholding (IHT) & Singular Value Projection (SVP)
- Gradient descent $x^{t+1/2} = x^t + \mu \mathbf{M}^T (\mathbf{M}x^t - y)$
- Projection $x^{t+1} = \text{prox}_{\Sigma}(x^{t+1/2})$
- Definition: proximal operator = projection onto the model Σ (similar to denoising)

$$\text{prox}_{\Sigma}(y) = \arg \min_{\tilde{x} \in \Sigma} \|y - \tilde{x}\|_2$$

NB: can be
NP-hard!

■ Theorem [Blumensath 2011]:

- PLA is instance optimal assuming \mathbf{M} has the RIP on $\Sigma - \Sigma$ with constants $\beta^2 \leq 1/\mu \leq 1.5\alpha^2$

corresponds to
 $\delta \leq 1/5$

Decoding by (convex?) optimization

■ RIP guarantees for generalized Basis Pursuit ?

■ General optimization framework

$$\Delta(y) := \arg \min_{x \in \mathcal{H}} f(x) \text{ s.t. } \|\mathbf{M}x - y\| \leq \epsilon$$

■ “One RIP to rule them all” [Traonmilin & G. 2015]



- Given cone (or Union of Subspace) model set Σ and regularizer $f(\cdot)$
- Definition of a constant $\delta_\Sigma(f)$

$$\delta_\Sigma(f) = \inf_{z \in \mathcal{T}_f(\Sigma)} \delta_\Sigma(z)$$

where $\delta_\Sigma(z) = \sup_{x \in \Sigma} \delta_\Sigma(x, z)$.

Expressions :

- $\delta_\Sigma^{UoS}(x, z) = \frac{-\mathcal{R}e\langle x, z \rangle}{\|x\|_{\mathcal{H}} \sqrt{\|x+z\|_{\Sigma}^2 - \|x\|_{\mathcal{H}}^2 - 2\mathcal{R}e\langle x, z \rangle}}$
- $\delta_\Sigma^{cone}(x, z) = \frac{-2\mathcal{R}e\langle x, z \rangle}{\|x+z\|_{\Sigma}^2 - 2\mathcal{R}e\langle x, z \rangle}$

Theorem

Assume that \mathbf{M} has the RIP on $\Sigma - \Sigma$ with constant $\delta < \delta_\Sigma(f)$ then the above decoder is instance optimal

Decoding by (convex?) optimization

■ RIP guarantees for generalized Basis Pursuit ?

■ General optimization framework

$$\Delta(y) := \arg \min_{x \in \mathcal{H}} f(x) \text{ s.t. } \|\mathbf{M}x - y\| \leq \epsilon$$

■ “One RIP to rule them all” [Traonmilin & G. 2015]

| | $f(x)$ | $\delta_\Sigma(f) \geq$ |
|-----------------------------------|---|--|
| ■ <i>sparsity</i> | $\ x\ _1$ | $1/\sqrt{2}$ |
| ■ <i>low-rank</i> | $\ x\ _\star$ | $1/\sqrt{J+2}$ |
| ■ <i>group sparsity</i> | $\ x\ _{1,2}$ | $1/\sqrt{J \frac{K_{\max}}{K_{\min}} + 2}$ |
| ■ <i>group-sparsity in levels</i> | $\sum_{j=1}^J \frac{\ x_j\ _{1,2}}{\sqrt{K_j}}$ | $2/3$ |
| ■ <i>permutation matrices</i> | $\ x\ _1$ | ... |
| ■ ... | Birkhoff polytope norm | ... |
| | ... | ... |