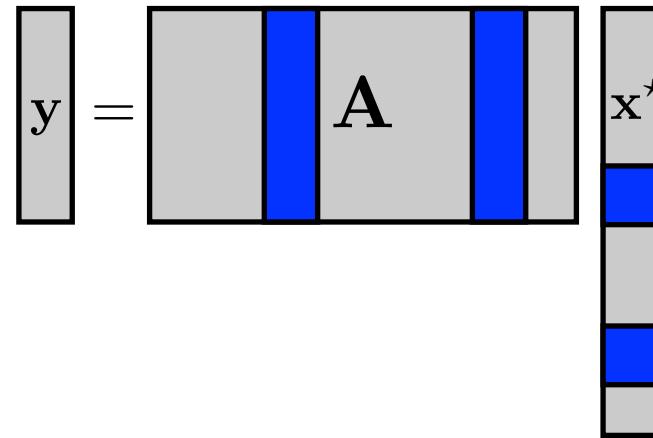


Sparse Representations with Partial Support Information

Cédric Herzet - INRIA Rennes

The sparsity has revealed to be a good prior in many inverse problems

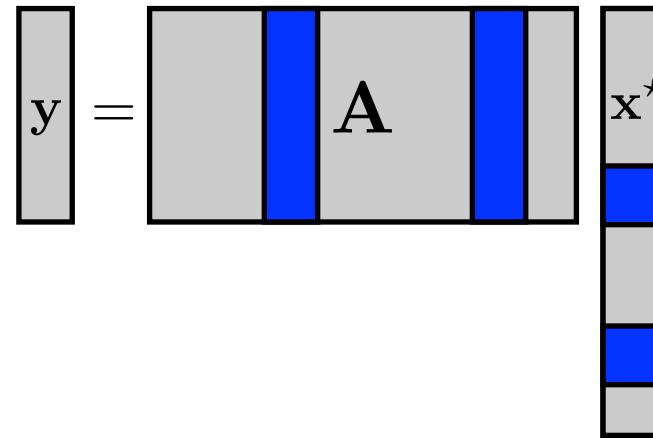
$$\mathbf{y} = \mathbf{A} \mathbf{x}^* \quad \mathcal{Q}^* = \{i \mid x_i^* \neq 0\}$$



$$\hat{\mathbf{x}} = \text{Algo}(\mathbf{y}, \text{sparsity})$$

Can we enhance the performance by
exploiting some support knowledge?

$$\mathbf{y} = \mathbf{A} \mathbf{x}^* + \text{"noise"} \quad \mathcal{Q}^* = \{i \mid x_i^* \neq 0\}$$



$$\mathcal{Q} = \mathcal{Q}^* + \text{"noise"}$$

$$\hat{\mathbf{x}} = \text{Algo}(\mathbf{y}, \text{sparsity}, \mathcal{Q})$$

Related material...

C. Soussen, R. Gribonval, J. Idier, C. Herzet, «*Joint k-step Analysis of Orthogonal Matching Pursuit and Orthogonal Least Squares*», IEEE Trans. Inf. Theory, May 2013.

C. Herzet , C. Soussen, J. Idier, R. Gribonval, «*Exact Recovery Conditions for Sparse Representations with Partial Support Information*», ArXiV 1305.4008

Image decomposition in wavelet domains

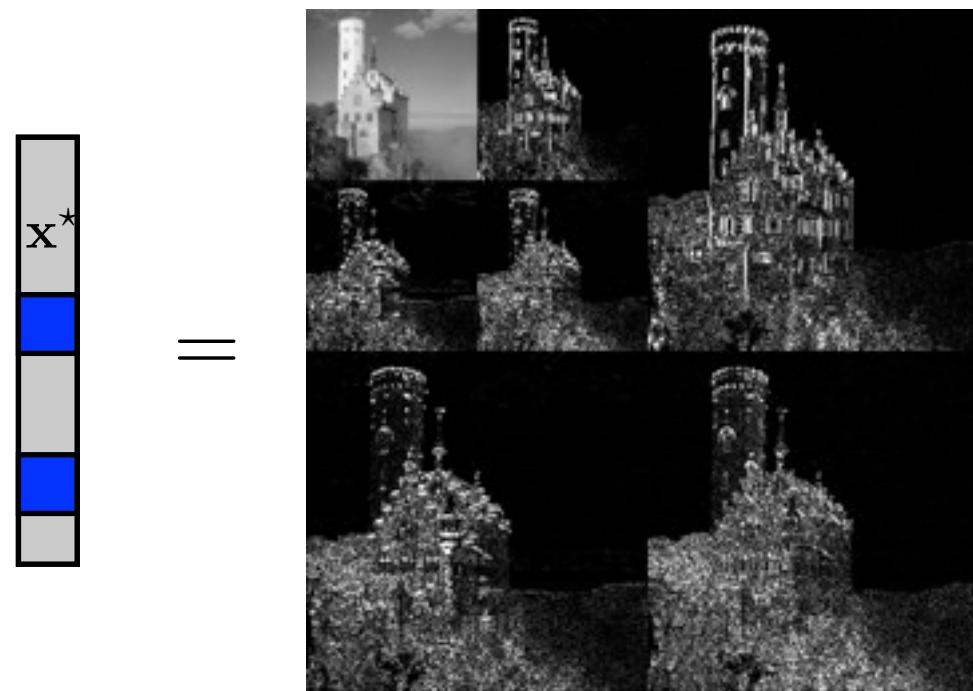
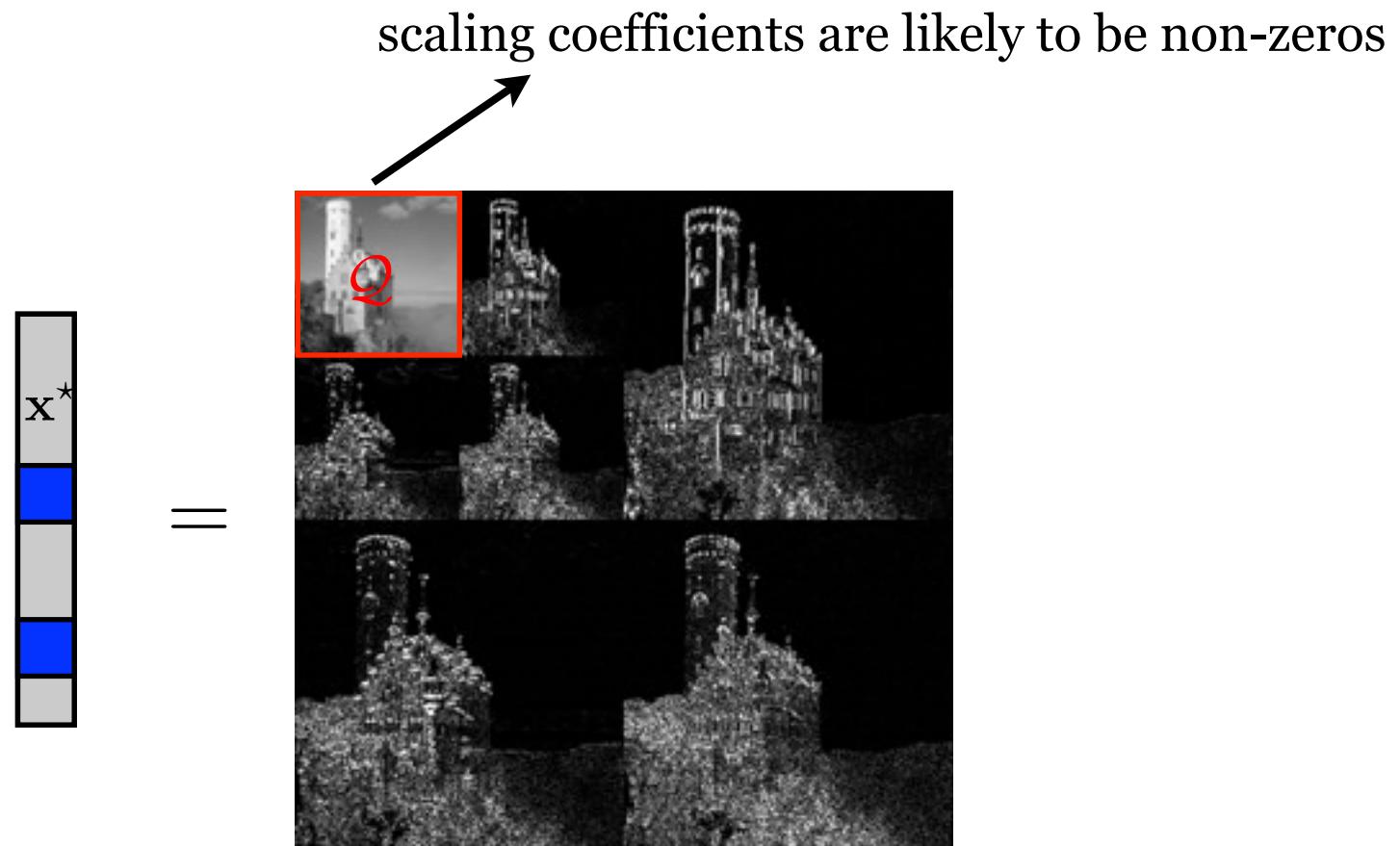
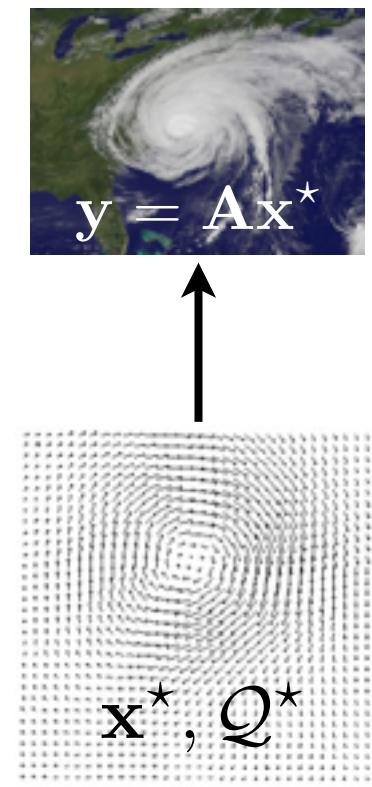


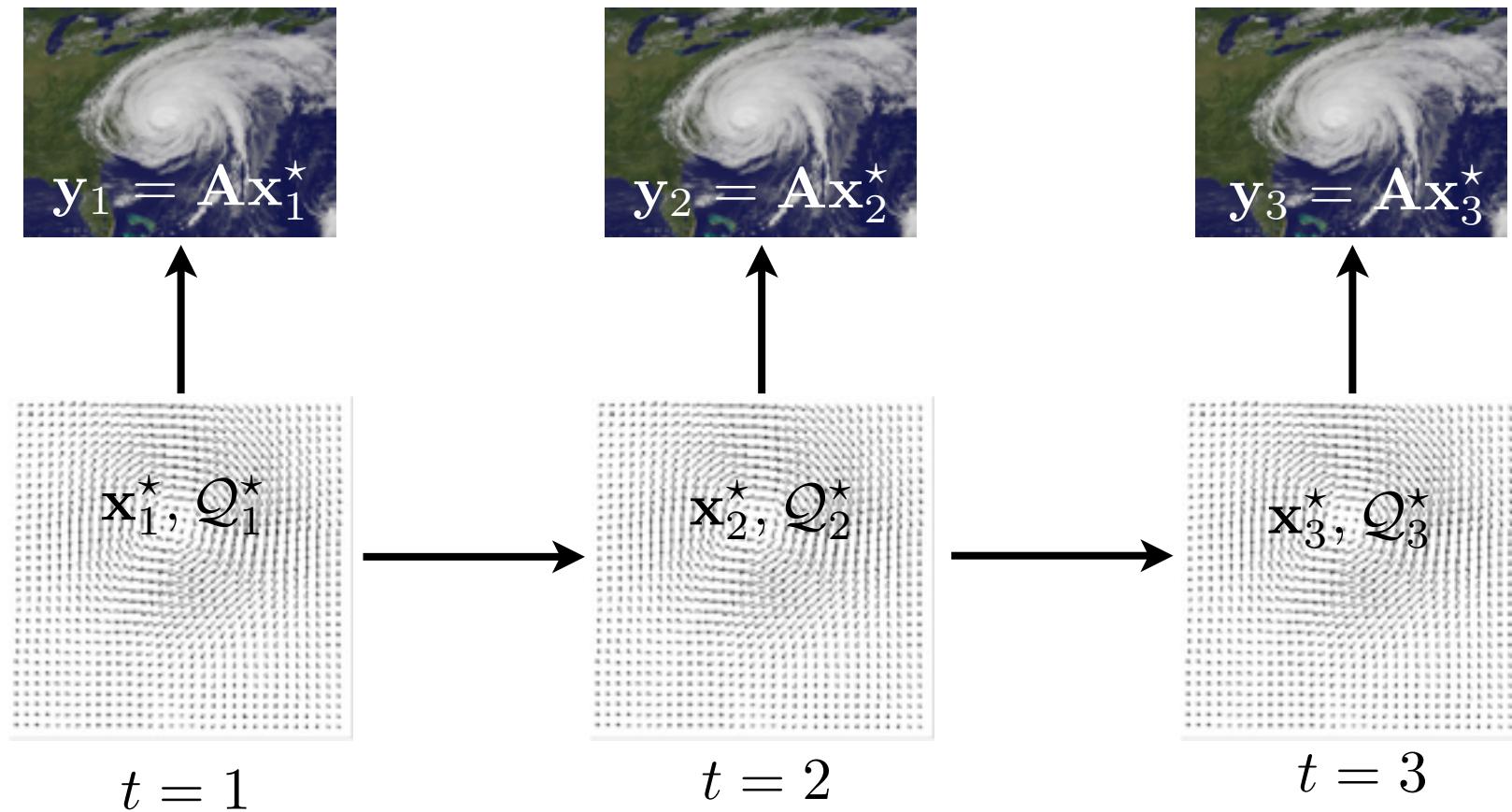
Image decomposition in wavelet domains



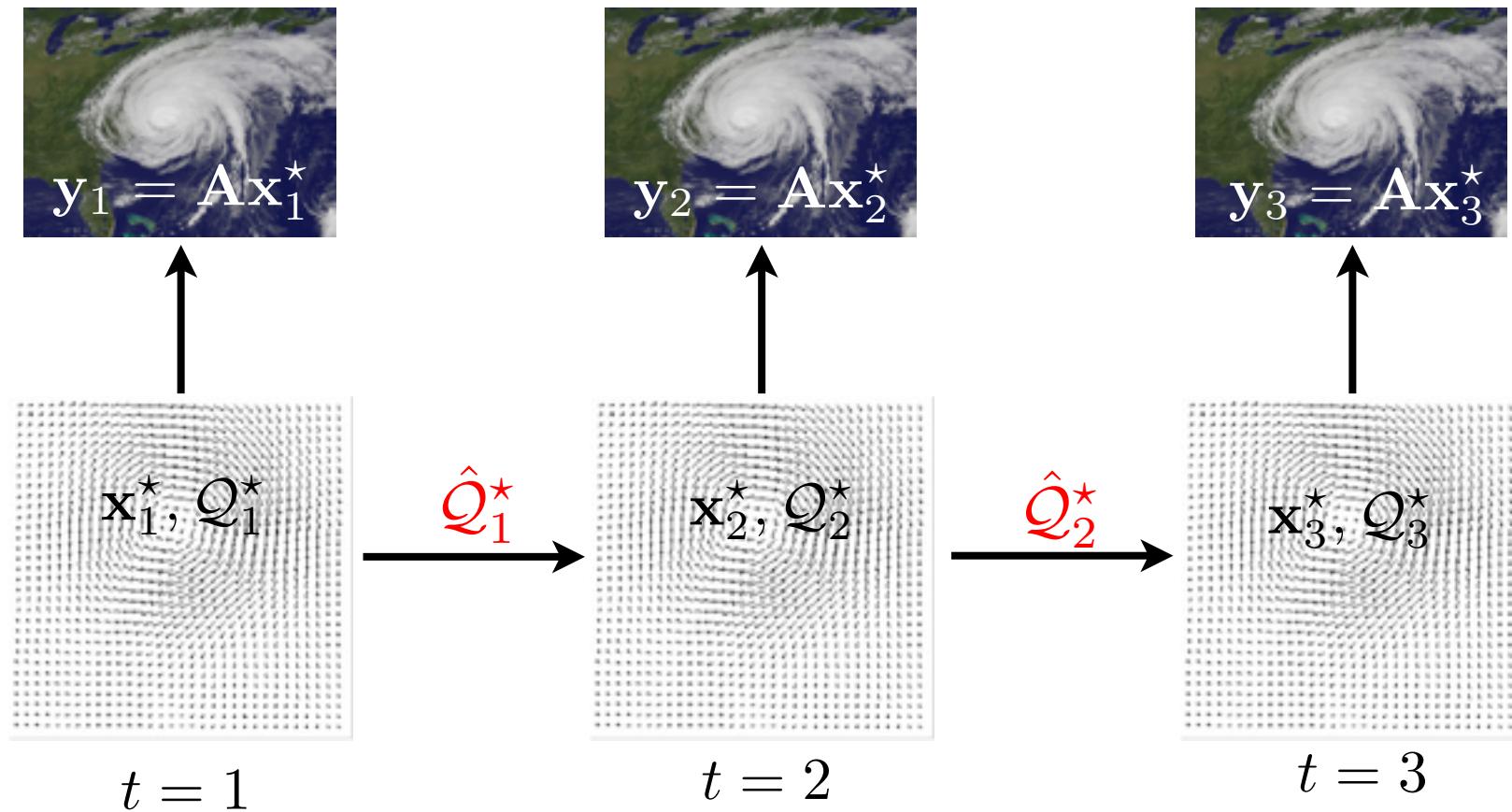
Sparse representations in slowly-varying dynamical models



Sparse representations in slowly-varying dynamical models



Sparse representations in slowly-varying dynamical models



$$\hat{\mathbf{x}}_i = \text{Algo}(\mathbf{y}_i, \text{sparsity}, \hat{\mathcal{Q}}_{i-1}^*)$$

Algorithms: how to exploit support information in practice...

The standard “ ℓ_0 ” sparse representation problem is of combinatorial nature

$$(P_0) : \quad \hat{\mathbf{x}} = \arg \min \|\mathbf{x}\|_0$$

subject to $\mathbf{y} = \mathbf{A}\mathbf{x}$

There exists a wide range of algorithms
for sparse representations

Greedy pursuit algorithms

Tackle the ℓ_0 problem via sequential minimizations

Procedures based on «norm relaxation»

Replace the ℓ_0 norm by ℓ_p norms

Bayesian methods

Define a prior promoting sparsity

There exists a wide range of algorithms for sparse representations

today's presentation

Greedy pursuit algorithms

Tackle the ℓ_0 problem via sequential minimizations



Procedures based on «norm relaxation»

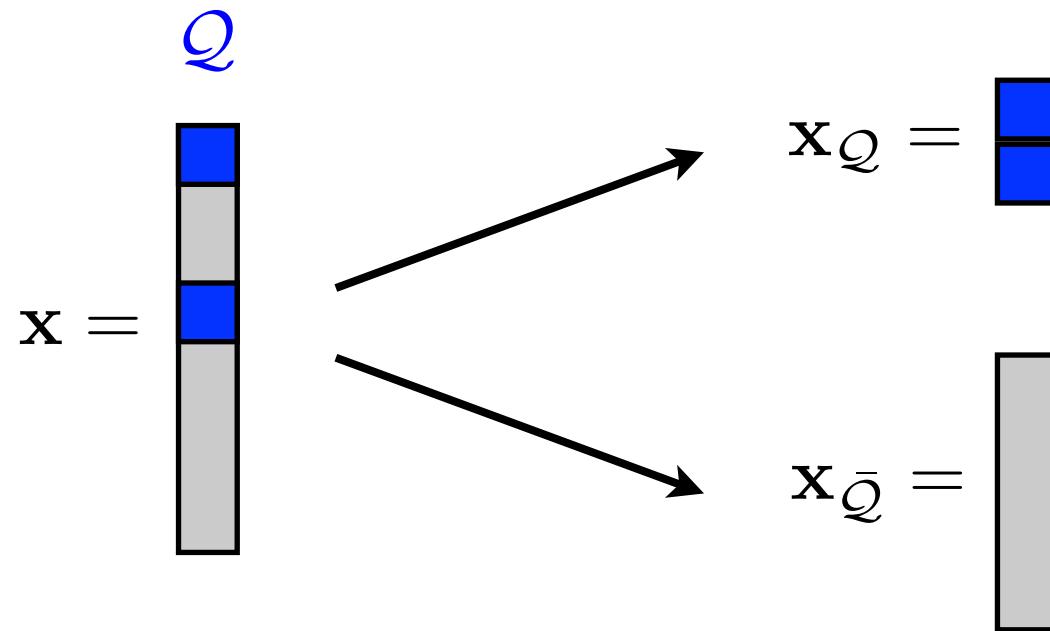
Replace the ℓ_0 norm by ℓ_p norms



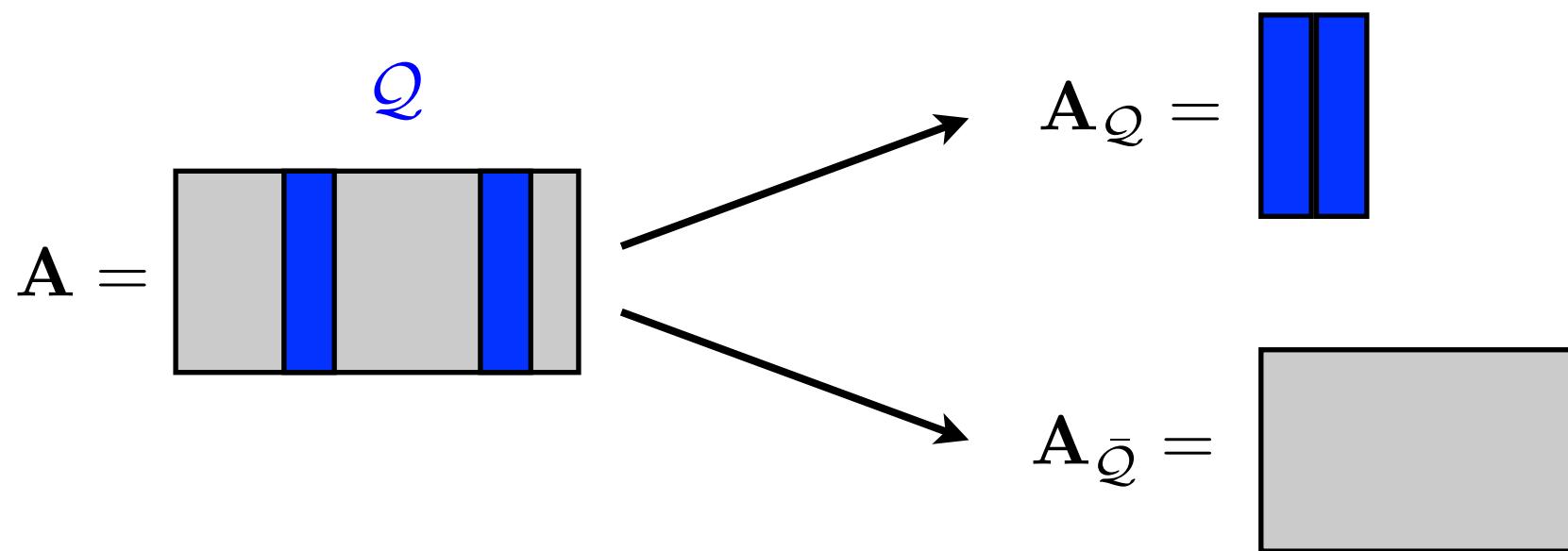
Bayesian methods

Define a prior promoting sparsity

Some useful notations: $\mathbf{x}_{\mathcal{Q}}$



Some useful notations: $\mathbf{A}_{\mathcal{Q}}$



Some useful notations: $\mathbf{P}_{\perp}^{\mathcal{Q}}$

$\mathbf{P}_{\perp}^{\mathcal{Q}}$: orthogonal projector onto $\text{span}(\mathbf{A}_{\mathcal{Q}})^{\perp}$

$\mathbf{v}^{\mathcal{Q}} \triangleq \mathbf{P}_{\perp}^{\mathcal{Q}} \mathbf{v}$: error for the best approximation of \mathbf{v}
in $\text{span}(\mathbf{A}_{\mathcal{Q}})$

Sparse algorithms in the standard setup: ℓ_p relaxation

$$(P_p) : \quad \hat{\mathbf{x}} = \arg \min \|\mathbf{x}\|_p \quad p \leq 1$$

subject to $\mathbf{y} = \mathbf{A}\mathbf{x}$

Sparse algorithms in the *informed* setup: ℓ_p relaxation

[Von Borries et al.2007], [Khajehnejad et al.2009], [Vaswani&Lu2010]

$$(P_{p,\mathcal{Q}}) : \quad \hat{\mathbf{x}} = \arg \min \|\mathbf{x}_{\bar{\mathcal{Q}}}\|_p \quad p \leq 1$$

subject to $\mathbf{y} = \mathbf{A}\mathbf{x}$

Sparse algorithms in the *standard* setup: orthogonal greedy algorithms

Init: $\hat{\mathcal{Q}}^{(0)} = \emptyset$

Repeat: $\hat{\mathcal{Q}}^{(l+1)} = \hat{\mathcal{Q}}^{(l)} \cup j$

with $j = \arg \min_i -|\mathbf{y}^T \mathbf{c}_i^{\hat{\mathcal{Q}}^{(l)}}|$

$$\mathbf{c}_i^{\hat{\mathcal{Q}}^{(l)}} = \begin{cases} \mathbf{a}_i^{\hat{\mathcal{Q}}^{(l)}} & (\text{OMP}) \\ \mathbf{a}_i^{\hat{\mathcal{Q}}^{(l)}} / \|\mathbf{a}_i^{\hat{\mathcal{Q}}^{(l)}}\| & (\text{OLS}) \end{cases}$$

Sparse algorithms in the *informed* setup: orthogonal greedy algorithms

Init: $\hat{\mathcal{Q}}^{(0)} = \mathcal{Q}$

Repeat: $\hat{\mathcal{Q}}^{(l+1)} = \hat{\mathcal{Q}}^{(l)} \cup j$

with $j = \arg \min_i -|\mathbf{y}^T \mathbf{c}_i^{\hat{\mathcal{Q}}^{(l)}}|$

$$\mathbf{c}_i^{\hat{\mathcal{Q}}^{(l)}} = \begin{cases} \mathbf{a}_i^{\hat{\mathcal{Q}}^{(l)}} & (\text{OMP}_{\mathcal{Q}}) \\ \mathbf{a}_i^{\hat{\mathcal{Q}}^{(l)}} / \|\mathbf{a}_i^{\hat{\mathcal{Q}}^{(l)}}\| & (\text{OLS}_{\mathcal{Q}}) \end{cases}$$

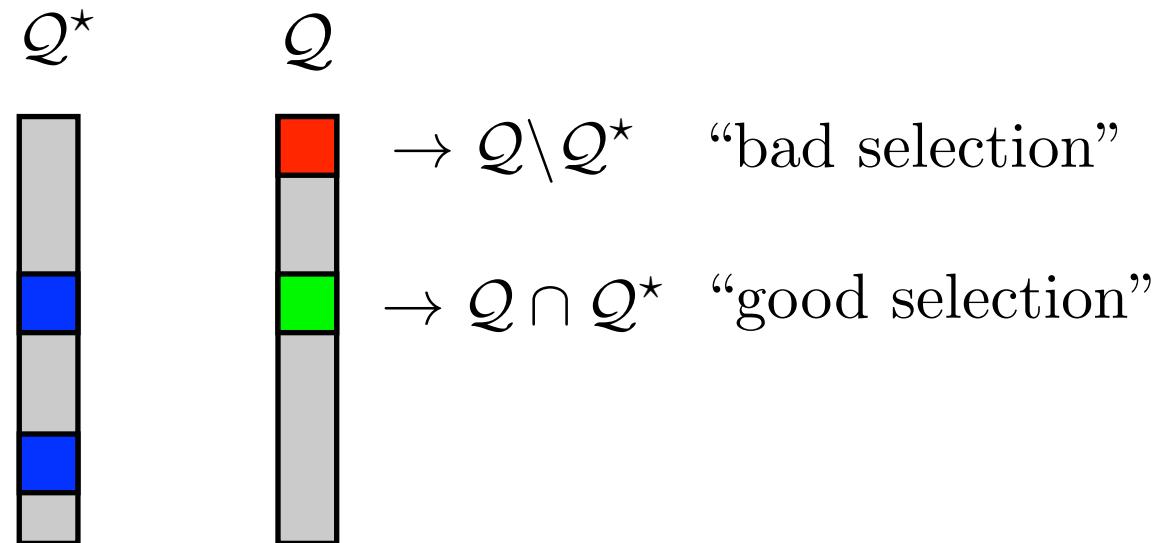
Guarantees of success in the informed setup

Under which conditions can one recover
the sought sparse vector?

$$\hat{\mathbf{x}} = \text{Algo}(\mathbf{y} = \mathbf{A}\mathbf{x}^*, \text{sparsity}, \mathcal{Q})$$

When is $\hat{\mathbf{x}} = \mathbf{x}^*$?

The support estimate may contain good and bad selections



$$\text{Card}\{\mathcal{Q}^*\} = k$$

$$\text{Card}\{\mathcal{Q} \cap \mathcal{Q}^*\} = g$$

$$\text{Card}\{\mathcal{Q} \setminus \mathcal{Q}^*\} = b$$

One can mainly distinguish between two types of conditions

Worst-case conditions

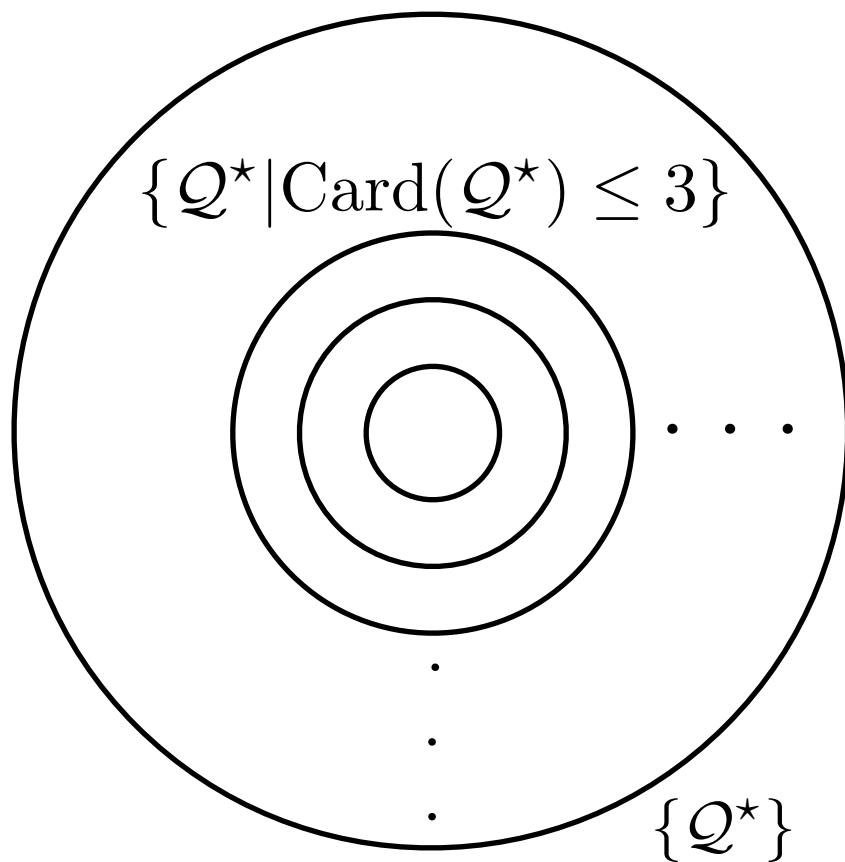
The sparse algorithm recovers the support for any amplitude $\mathbf{x}_{\mathcal{Q}^}^*$*



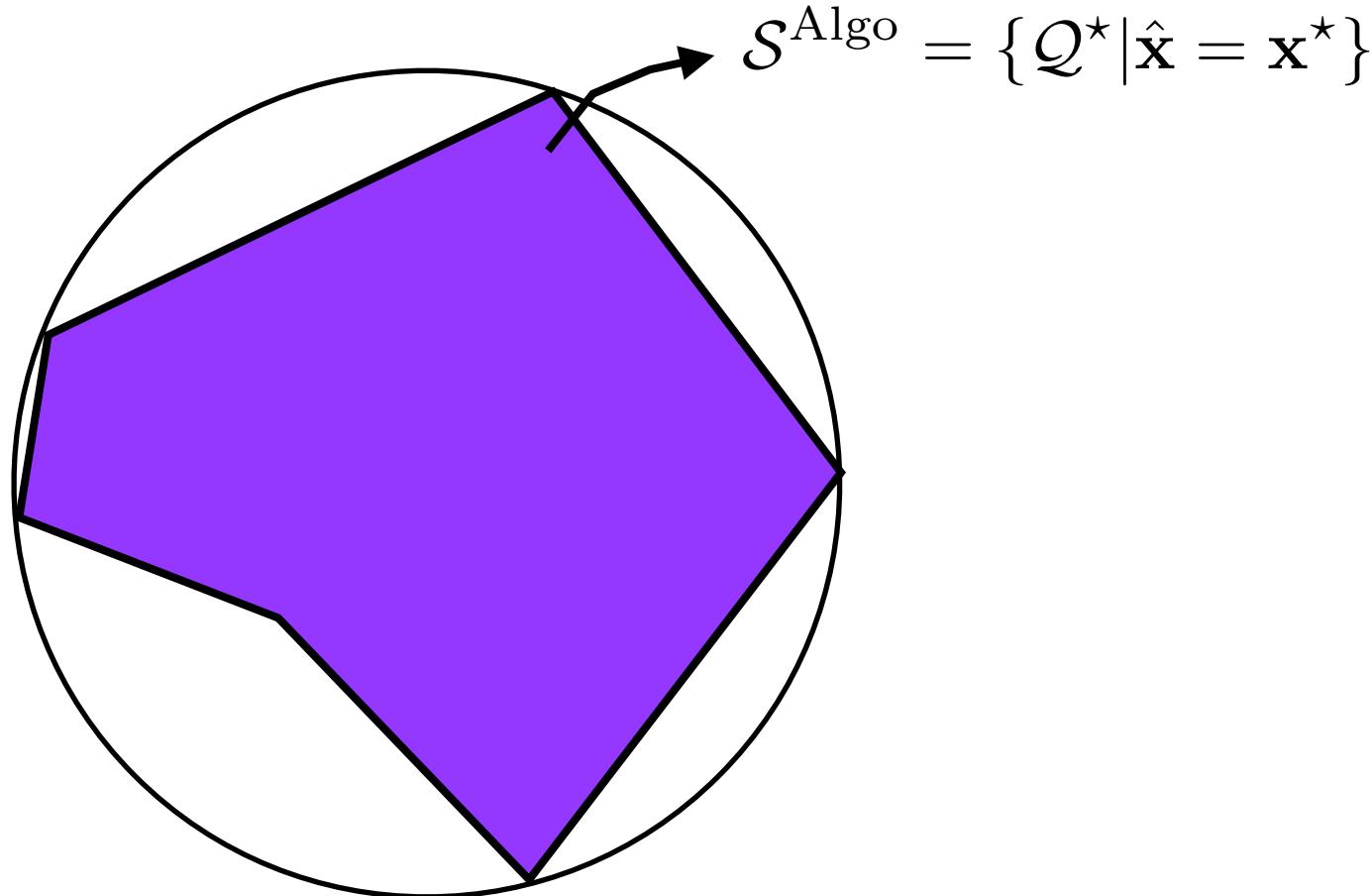
Conditions in probability

The sparse algorithm recovers the support for «most» $\mathbf{x}_{\mathcal{Q}^}^*$'s*

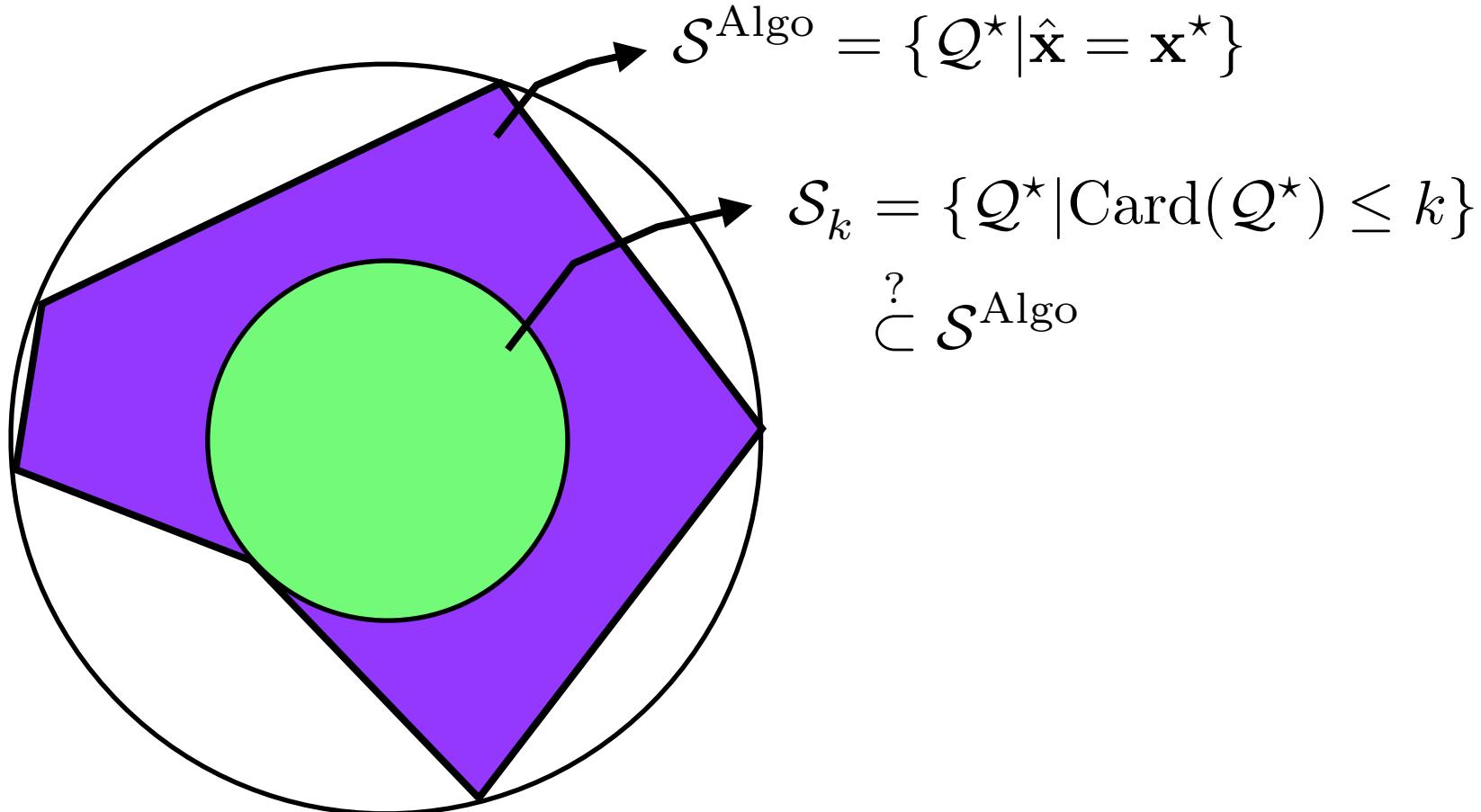
Visualization of sparsity-based recovery conditions



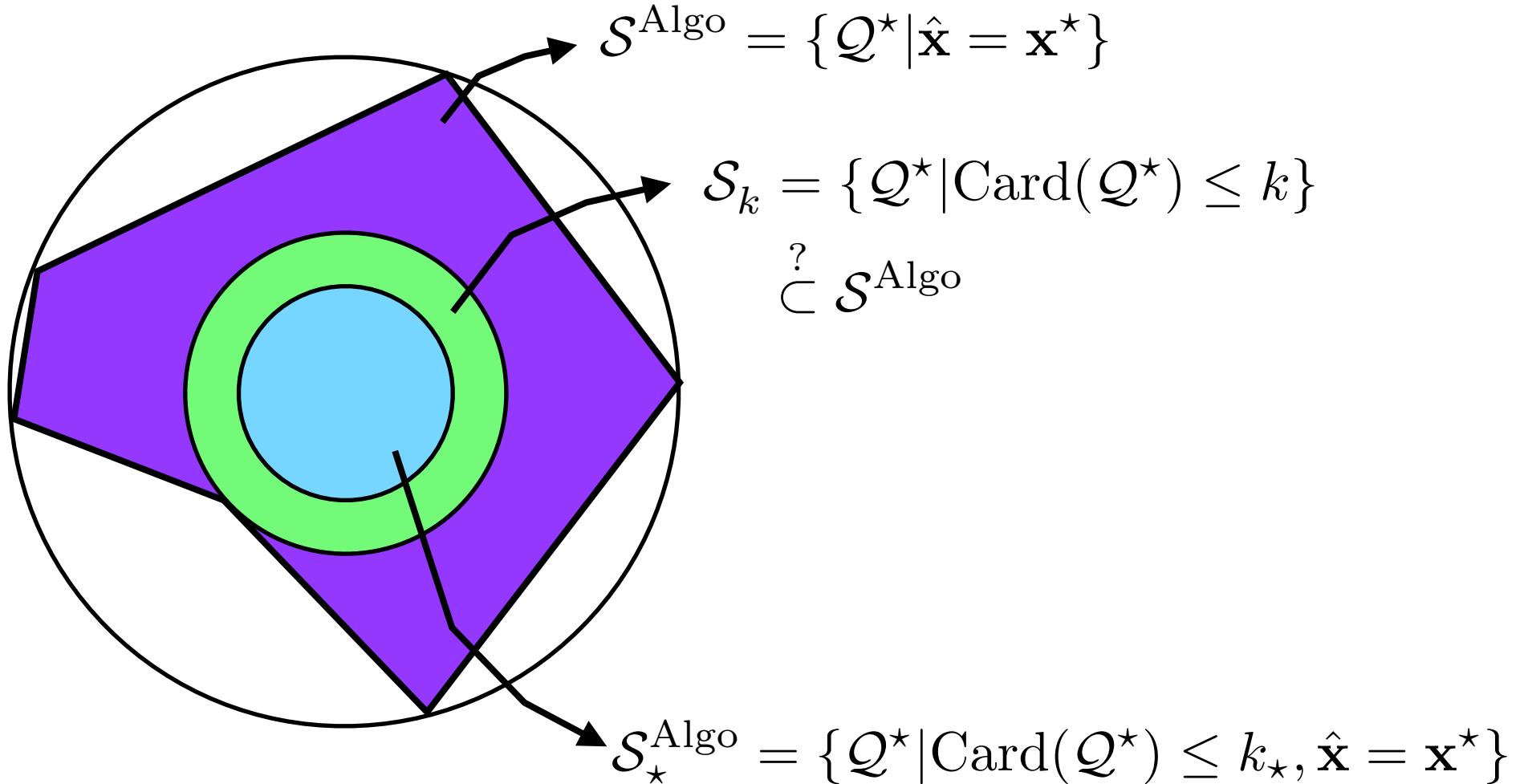
Visualization of sparsity-based recovery conditions



Visualization of sparsity-based recovery conditions



Visualization of sparsity-based recovery conditions



Mutual coherence and RIC: two characterizations of the dictionary

Mutual coherence:

$$\mu \triangleq \max_{i \neq j} |\mathbf{a}_i^T \mathbf{a}_j|$$

k -th order restricted isometry constants:

$$(1 - \delta_k) \|\mathbf{x}\|^2 \leq \|\mathbf{Ax}\|^2 \leq (1 + \delta_k) \|\mathbf{x}\|^2$$

for any k -sparse vector \mathbf{x} .

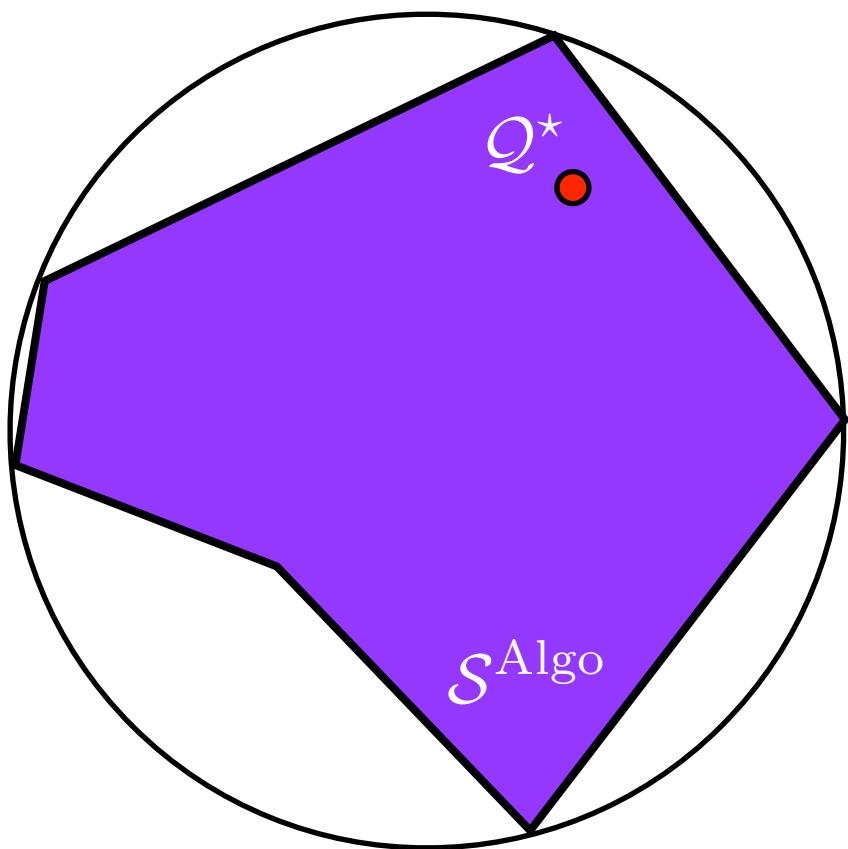
We address three main questions:

Can one easily characterize $\mathcal{S}^{\text{Algo}}$, $\mathcal{S}_*^{\text{Algo}}$?

How tight are the conditions defining $\mathcal{S}_*^{\text{Algo}}$?

Implication between the success of different algorithms ?

Characterization of $\mathcal{S}^{\text{Algo}}$ in the standard setup



$Q^* \in \mathcal{S}^{\text{Algo}}$ iff:

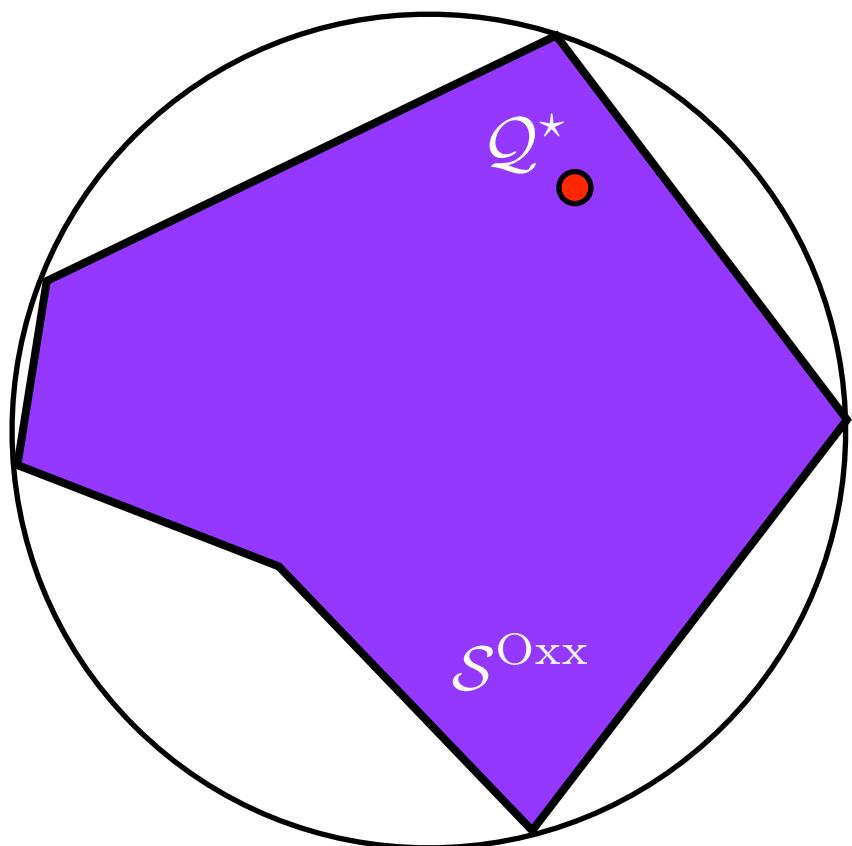
(P_p) : [Gribonval & Nielsen 2003]

$$\text{NSP}_p : \max_{\mathbf{v} \in \ker_0(\mathbf{A})} \frac{\|\mathbf{v}_{Q^*}\|_p}{\|\mathbf{v}_{\bar{Q}^*}\|_p} < 1$$

Oxx : [Tropp 2004], [Soussen et al. 2013]

$$\text{ERC} : \max_{i \notin Q^*} \|\mathbf{A}_{Q^*}^+ \mathbf{a}_i\|_1 < 1$$

Characterization of $\mathcal{S}^{\text{Oxx}_{\mathcal{Q}}}$ in the informed setup



$Q^* \in \mathcal{S}^{\text{Algo}}$ iff:

Oxx : [Tropp 2004], [Soussen et al. 2013]

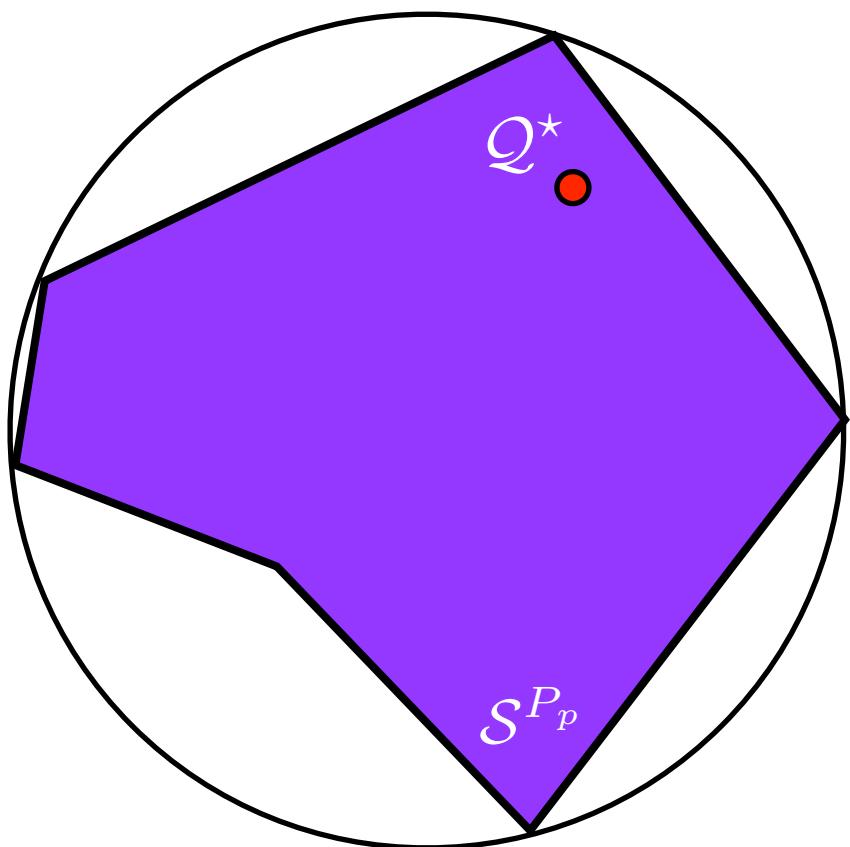
$$\text{ERC} : \max_{i \notin \mathcal{Q}^*} \|\mathbf{A}_{\mathcal{Q}^*}^+ \mathbf{a}_i\|_1 < 1$$

$\text{Oxx}_{\mathcal{Q}} : [\text{Soussen et al. 2013}]$

ERC_Oxx $_{\mathcal{Q}}$:

$$\max_{i \notin \mathcal{Q}^*} \|(\mathbf{C}_{\mathcal{Q}^* \setminus \mathcal{Q}}^{\mathcal{Q}})^+ \mathbf{c}_i^{\mathcal{Q}}\|_1 < 1$$

Characterization of $\mathcal{S}^{P_p, \mathcal{Q}}$ in the informed setup



$Q^* \in \mathcal{S}^{\text{Algo}}$ iff:

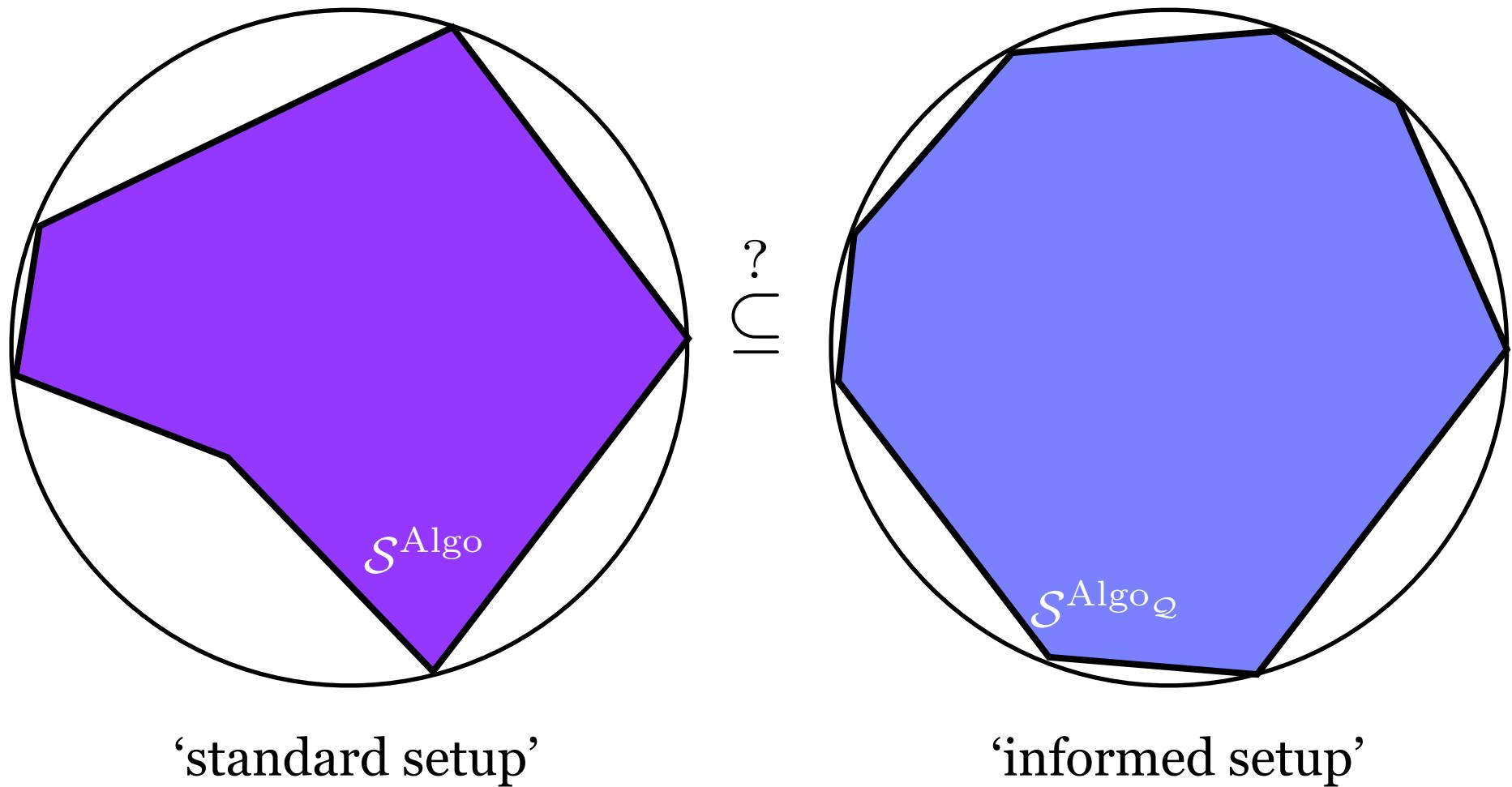
(P_p) : [Gribonval & Nielsen 2003]

$$\text{NSP}_p : \max_{\mathbf{v} \in \ker_0(\mathbf{A})} \frac{\|\mathbf{v}_{\mathcal{Q}^*}\|_p}{\|\mathbf{v}_{\bar{\mathcal{Q}}^*}\|_p} < 1$$

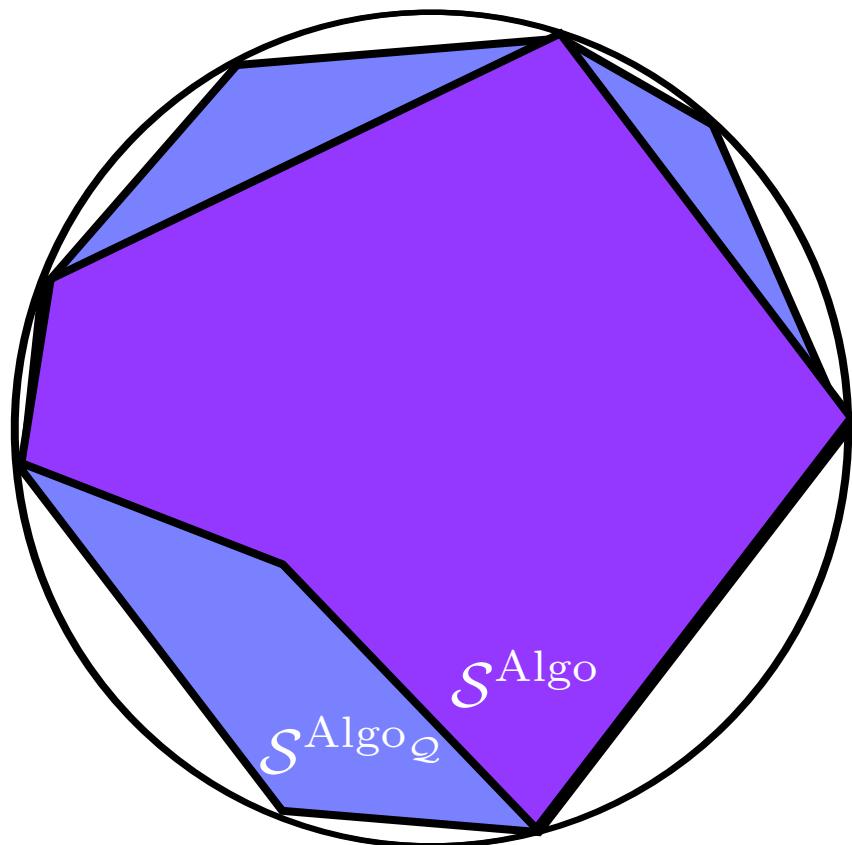
$(P_{p, \mathcal{Q}})$: [Wang & Yin 2010], [Chen & Zou 2010],
[Bandeira et al. 2011]

$$\text{NSP}_p(\mathcal{Q}) : \max_{\mathbf{v} \in \ker_0(\mathbf{A})} \frac{\|\mathbf{v}_{\mathcal{Q}^* \setminus \mathcal{Q}}\|_p}{\|\mathbf{v}_{\overline{\mathcal{Q}^* \cup \mathcal{Q}}}\|_p} < 1$$

A «good» support estimate
weakens the conditions of success



A «good» support estimate weakens the conditions of success



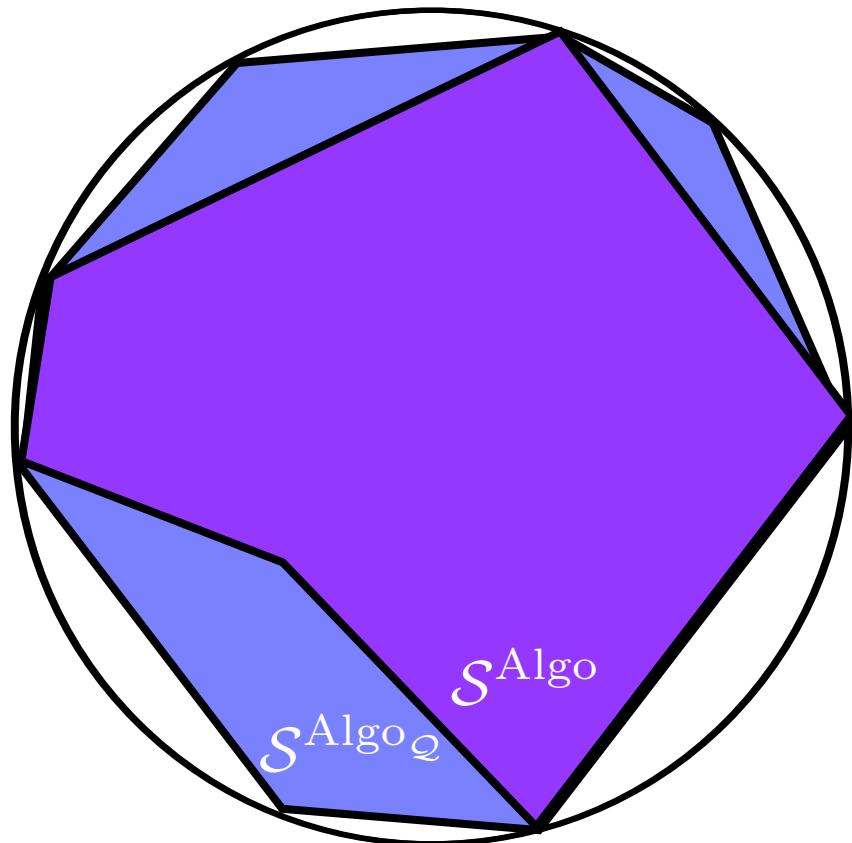
if $\mathcal{Q} \subset \mathcal{Q}^*$:

$$\mathcal{S}^{\text{Algo}} \subseteq \mathcal{S}^{\text{Algo}_{\mathcal{Q}}}$$

$(P_{p,\mathcal{Q}})$: [Bandeira et al. 2011]

$\text{Oxx}_{\mathcal{Q}}$: [Soussen et al. 2013]

A «good» support estimate weakens the conditions of success



if $\mathcal{Q} \subset \mathcal{Q}^*$:

$$\mathcal{S}^{\text{Algo}} \subseteq \mathcal{S}^{\text{Algo}_\mathcal{Q}}$$

$(P_{p,\mathcal{Q}})$: [Bandeira et al. 2011]

$\text{Oxx}_\mathcal{Q}$: [Soussen et al. 2013]

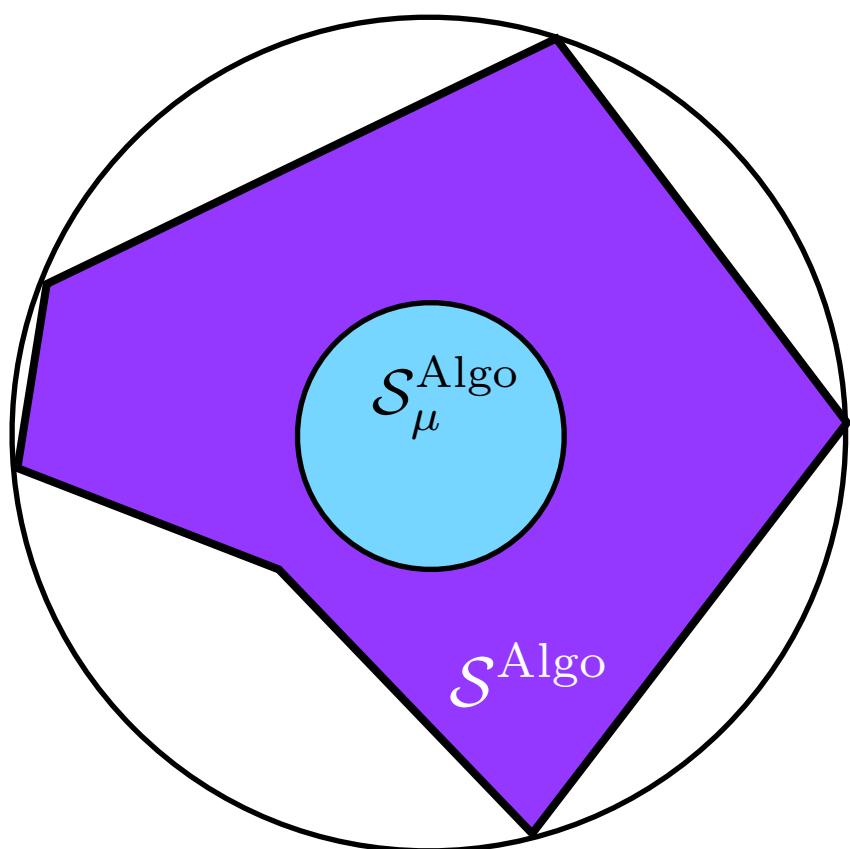
Quid for $\mathcal{Q} \not\subseteq \mathcal{Q}^*$?

Better guarantee for $P_{0,Q}$ as soon as
 Q contains at least 50% of «good» atoms

$$\mathcal{S}_k \subseteq \mathcal{S}^{P_0} \quad \text{iff} \quad k < \frac{1}{2}\text{spark}(\mathbf{A})$$

$$\mathcal{S}_k \subseteq \mathcal{S}^{P_0,Q} \quad \text{iff} \quad k < \frac{1}{2}(\text{spark}(\mathbf{A}) + g - b) \quad (k + b < \text{spark}(\mathbf{A}))$$

Characterization of $\mathcal{S}_\mu^{\text{Algo}}$



$Q^* \in \mathcal{S}_\mu^{\text{Algo}}$ iff:

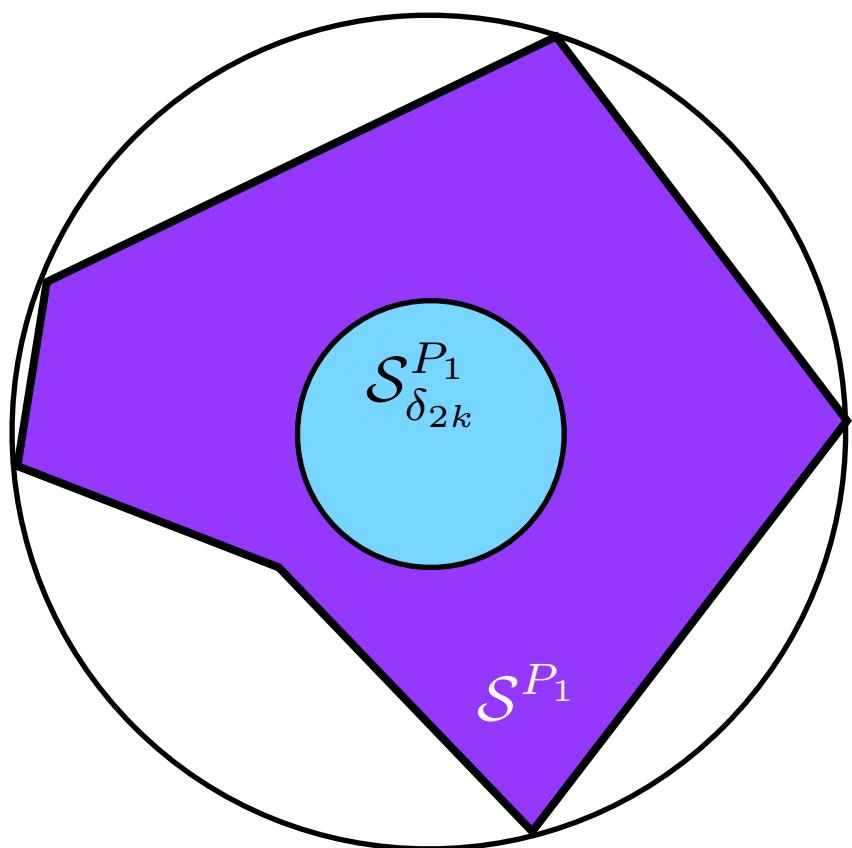
All : [Gribonval & Nielsen 2003], [Fuchs 2004]
[Tropp 2004]

$$\mu < \frac{1}{2k - 1}$$

All Q : [Herzet et al. 2013]

$$\mu < \frac{1}{2k - g + b - 1}$$

Characterization of $\mathcal{S}_{\delta_{2k}}^{P_1}$



$\mathcal{Q}^* \in \mathcal{S}_{\delta_{2k}}^{\text{Algo}}$ iff:

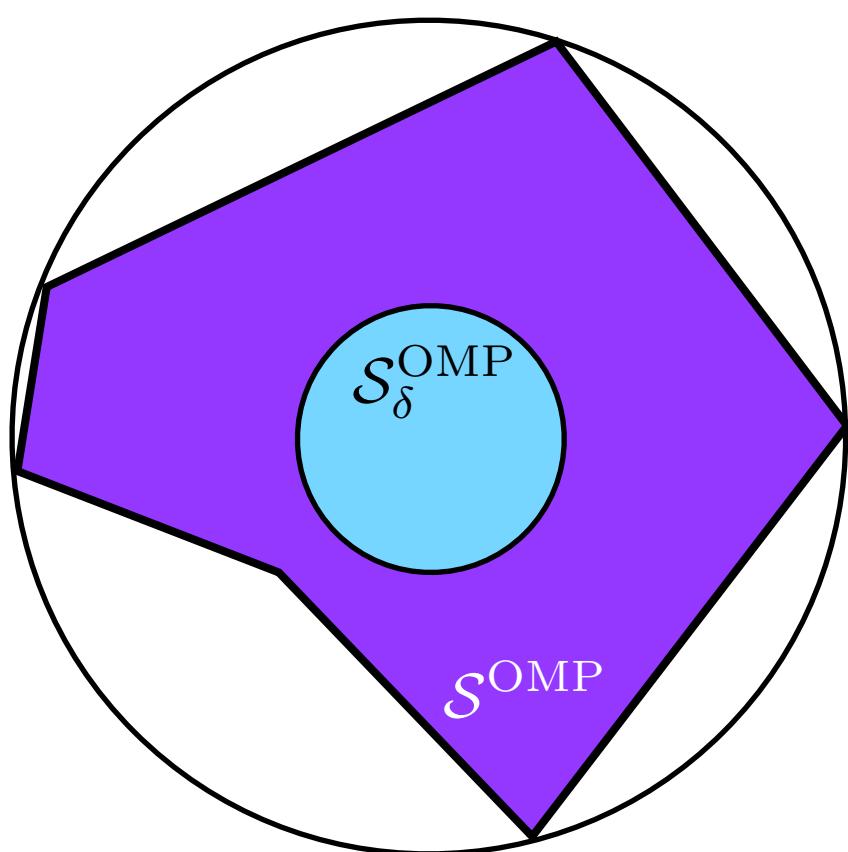
$(P_1) : [Candes2008]$

$$\delta_{2k} < \left(1 + \sqrt{2}\right)^{-1}$$

$(P_{1,\mathcal{Q}}) : [Friedlander et al.2012]$

$$\delta_{2k} < \left(1 + \sqrt{2 \left(1 + \frac{b-g}{k}\right)}\right)^{-1}$$

Characterization of $\mathcal{S}_\delta^{\text{OMP}}$



$\mathcal{Q}^* \in \mathcal{S}_\delta^{\text{Algo}}$ iff:

Oxx : [Maleh2011], [Mo&Shen2012]
[Wang&Shim2012]

$$\delta_{k+1} < \frac{1}{\sqrt{k} + 1}$$

OMP $_{\mathcal{Q}}$: [Karahanoglu&Erdogna2012]

$$\delta_{k+b+1} < \frac{1}{\sqrt{k-g} + 1}$$

We address three main questions:

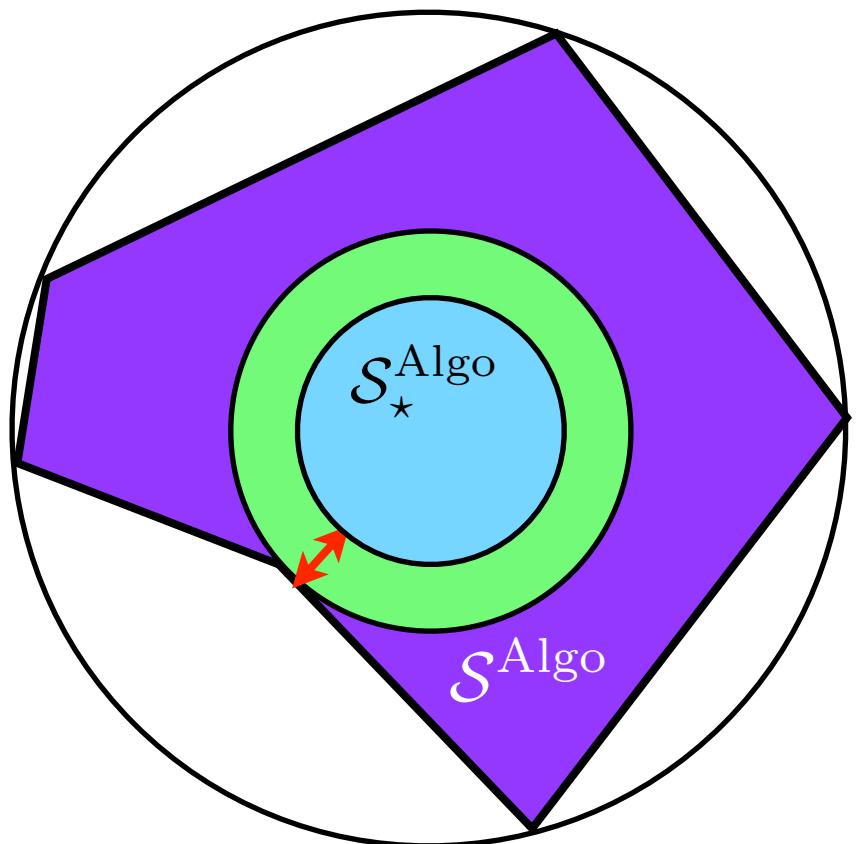
Can one easily characterize $\mathcal{S}^{\text{Algo}}$, $\mathcal{S}_\star^{\text{Algo}}$?

The conditions in the standard setup can be generalized and weakened if the support estimate is reliable.

How tight are the conditions defining $\mathcal{S}_\star^{\text{Algo}}$?

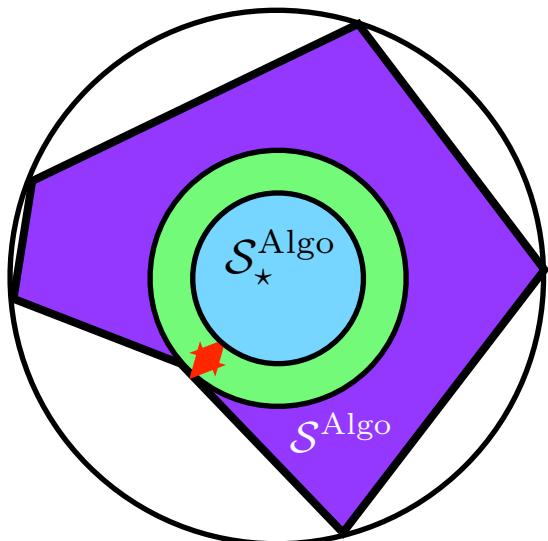
Implication between the success of different algorithms ?

Is there a dictionary
for which $\mathcal{S}_\star^{\text{Algo}}$ is tight?

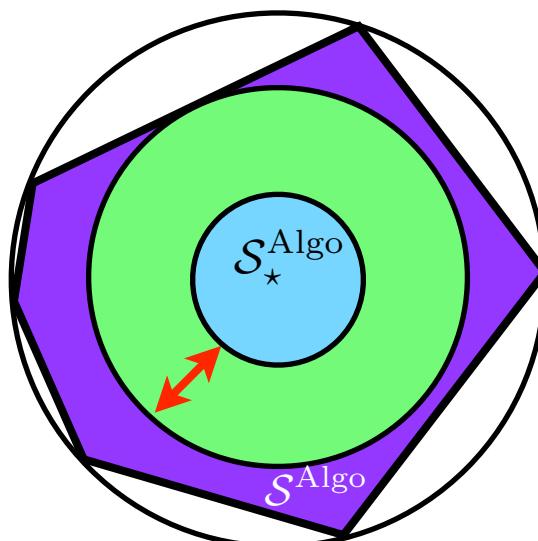


Is there a dictionary
for which $\mathcal{S}_\star^{\text{Algo}}$ is tight?

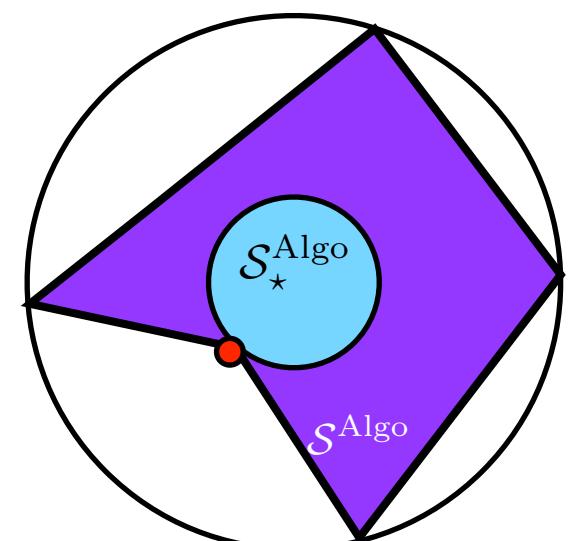
$$\mu(\mathbf{A}_i) = \mu \quad \forall i$$



\mathbf{A}_1



\mathbf{A}_2



\mathbf{A}_3

Tightness of the different conditions

$$\mu < \frac{1}{2k - g + b - 1} : \text{tight for } \text{Oxx}_{\mathcal{Q}} \text{ and } (P_{p, \mathcal{Q}})$$

[Herzet et al. 2013]

$$\delta_{k+b+1} < \frac{1}{\sqrt{k-g} + 1} : \text{«quasi-tight» for } \text{Oxx}_{\mathcal{Q}} \quad (\delta_{k+b+1} = \frac{1}{\sqrt{k-g}})$$

[Herzet et al. 2013]

We address three main questions:

Can one easily characterize $\mathcal{S}^{\text{Algo}}, \mathcal{S}_k^{\text{Algo}}, \mathcal{S}_\mu^{\text{Algo}}$?

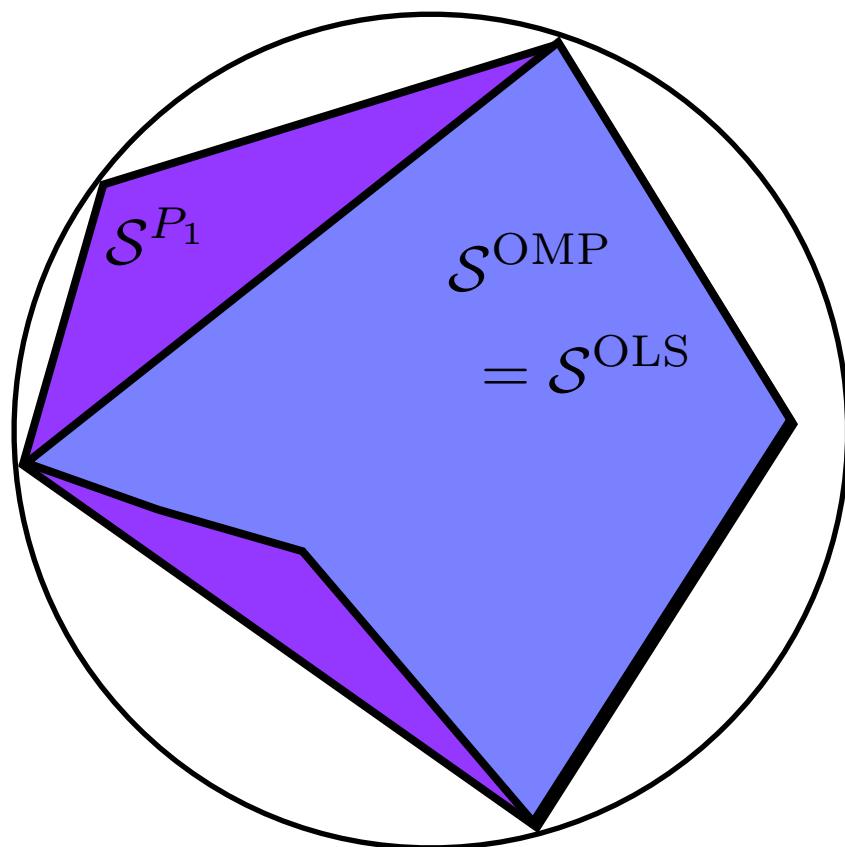
The conditions in the standard setup can be generalized and weakened if the support estimate is reliable.

How tight are the conditions defining $\mathcal{S}_\mu^{\text{Algo}}$?

The coherence condition is tight for all the algorithms. The RIC condition is quasi-tight for Oxx.

Implication between the success of different algorithms ?

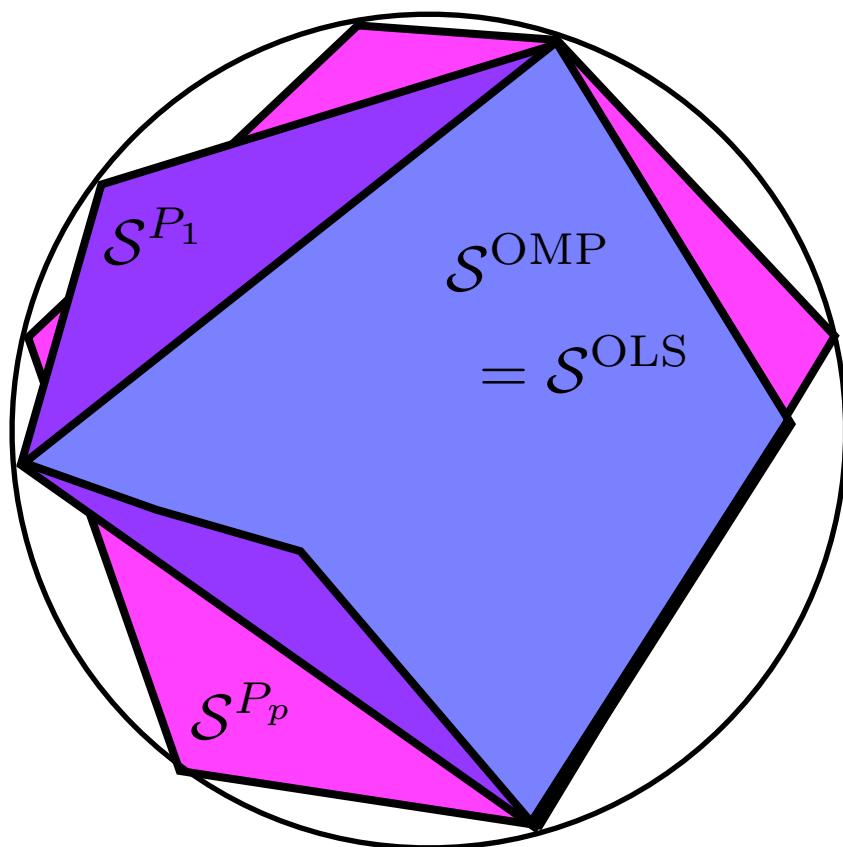
Inclusions in the standard setup...



$$1) \mathcal{S}^{\text{Oxx}} \subseteq \mathcal{S}^{P_1}$$

[Tropp 2004]

Inclusions in the standard setup...

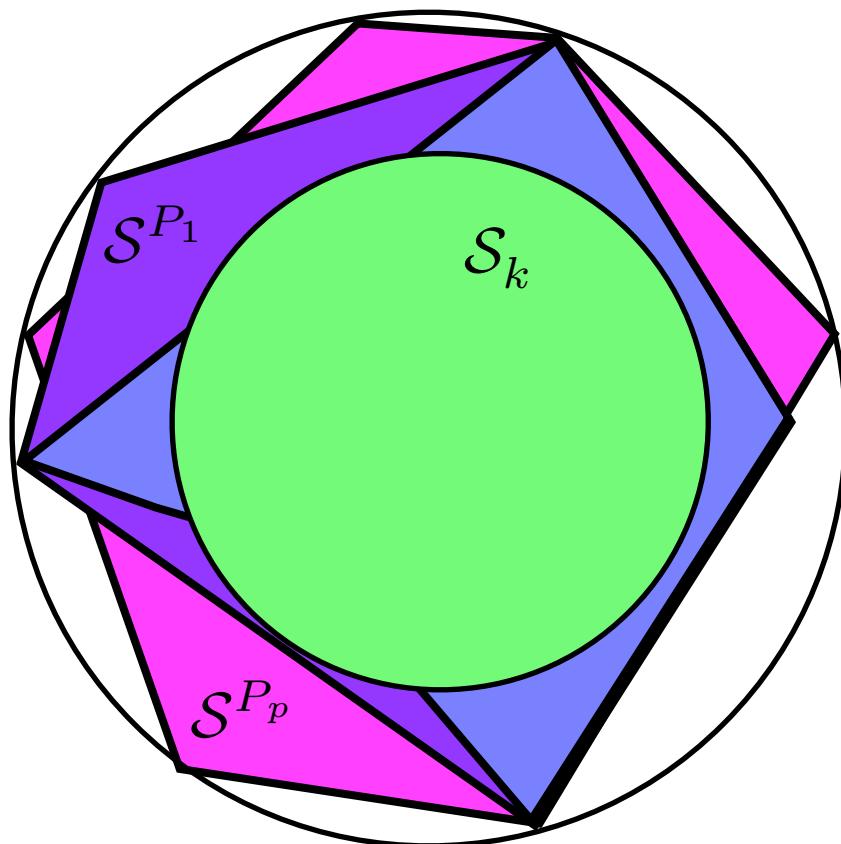


$$1) \mathcal{S}^{\text{Oxx}} \subseteq \mathcal{S}^{P_1}$$

[Tropp 2004]

$$2) \mathcal{S}^{P_p} \not\subseteq \mathcal{S}^{P_q} \quad p \neq q$$

Inclusions in the standard setup...



$$1) \mathcal{S}^{\text{Oxx}} \subseteq \mathcal{S}^{P_1} \quad [\text{Tropp 2004}]$$

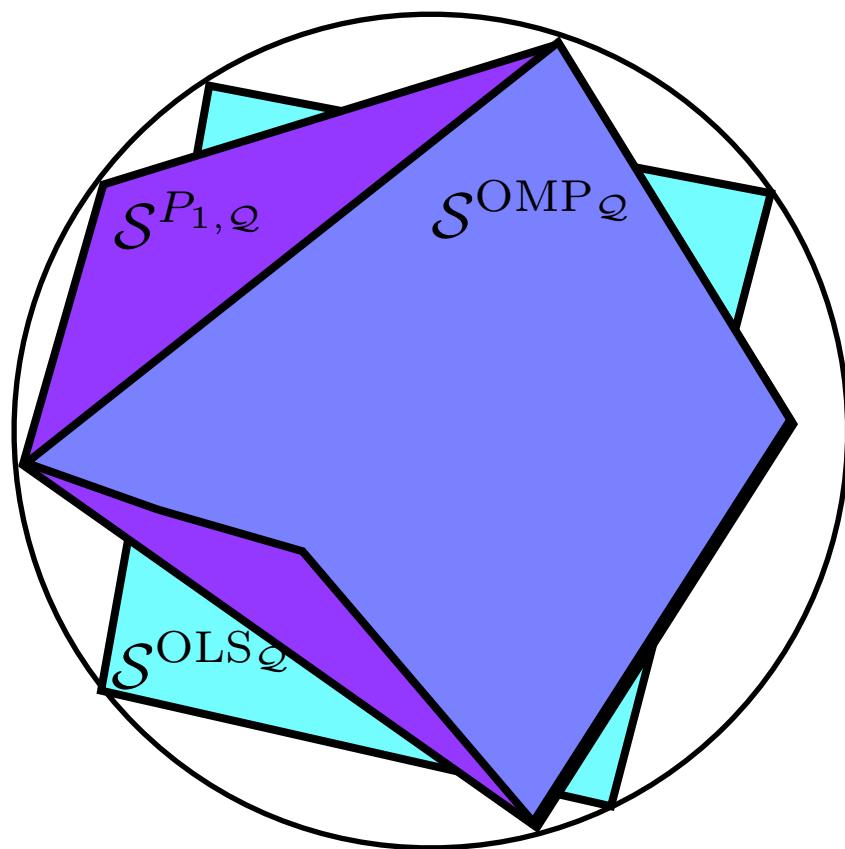
$$2) \mathcal{S}^{P_p} \not\subseteq \mathcal{S}^{P_q} \quad p \neq q$$

$$3) \mathcal{S}_k \subseteq \mathcal{S}^{P_p} \stackrel{q \leq p}{\Leftrightarrow} \mathcal{S}_k \subseteq \mathcal{S}^{P_q}$$

[Gribonval & Nielsen 2007]

Inclusions in the informed setup...

[Herzet et al.2013] [Souussen et al.2013]



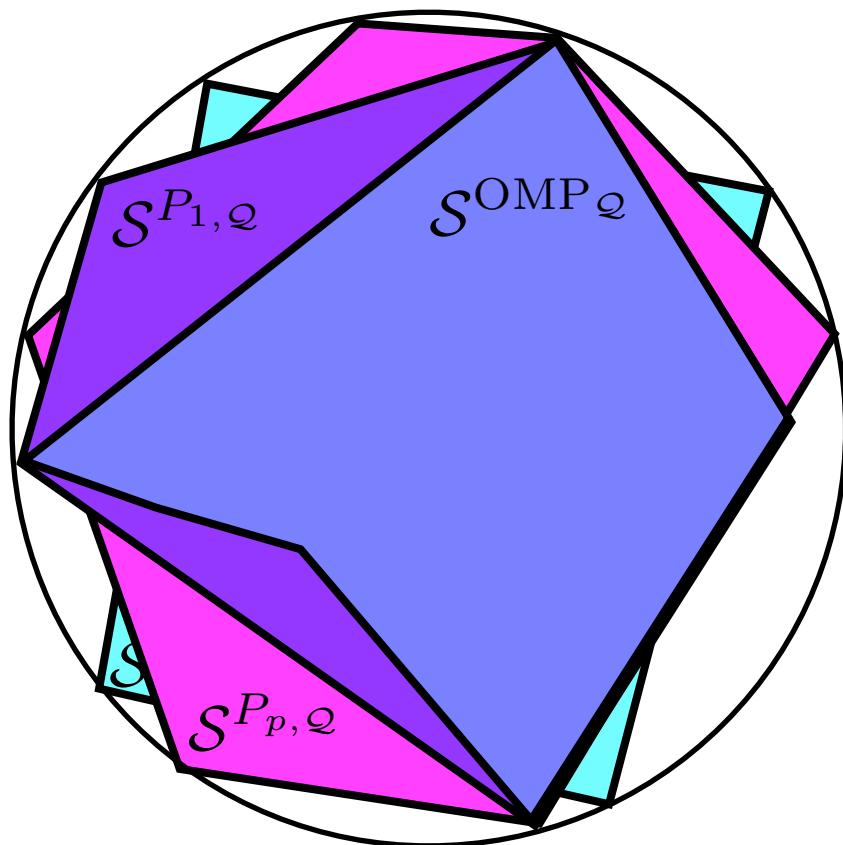
1) $\mathcal{S}^{\text{OMP}_Q} \not\subseteq \mathcal{S}^{\text{OLS}_Q}$

$$\mathcal{S}^{\text{OMP}_Q} \subseteq \mathcal{S}^{P_1,Q}$$

$\mathcal{S}^{\text{OLS}_Q} \not\subseteq \mathcal{S}^{P_1,Q}$

Inclusions in the informed setup...

[Herzet et al.2013] [Soussen et al.2013]



1) $\mathcal{S}^{\text{OMP}_Q} \not\subseteq \mathcal{S}^{\text{OLS}_Q}$

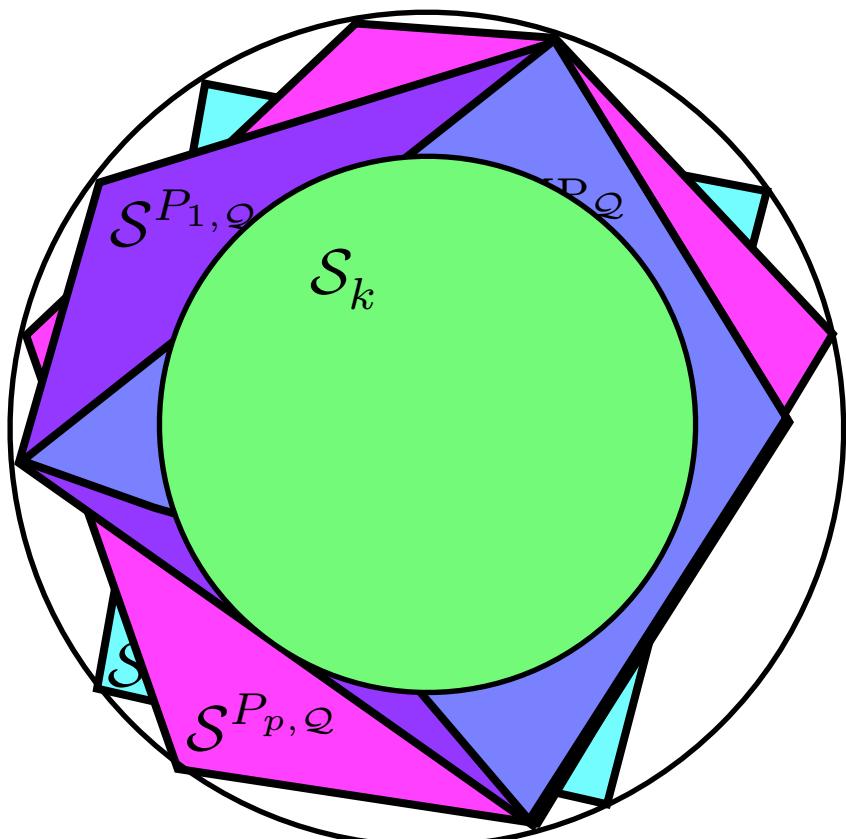
$$\mathcal{S}^{\text{OMP}_Q} \subseteq \mathcal{S}^{P_1,Q}$$

$$\mathcal{S}^{\text{OLS}_Q} \not\subseteq \mathcal{S}^{P_1,Q}$$

2) $\mathcal{S}^{P_p,Q} \not\subseteq \mathcal{S}^{P_q,Q} \quad p \neq q$

Inclusions in the informed setup...

[Herzet et al.2013] [Souussen et al.2013]



1) $\mathcal{S}^{\text{OMP}_{\mathcal{Q}}} \not\subseteq \mathcal{S}^{\text{OLS}_{\mathcal{Q}}}$

$$\mathcal{S}^{\text{OMP}_{\mathcal{Q}}} \subseteq \mathcal{S}^{P_1, \mathcal{Q}}$$

$\mathcal{S}^{\text{OLS}_{\mathcal{Q}}} \not\subseteq \mathcal{S}^{P_1, \mathcal{Q}}$

2) $\mathcal{S}^{P_p, \mathcal{Q}} \not\subseteq \mathcal{S}^{P_q, \mathcal{Q}} \quad p \neq q$

3) $\mathcal{S}_k \subset \bigcap_{\substack{\text{Card}(\mathcal{Q}^* \cap \mathcal{Q}) = g \\ \text{Card}(\mathcal{Q} \setminus \mathcal{Q}^*) = b}} \mathcal{S}^{P_p, \mathcal{Q}}$

$\stackrel{q \leq p}{\Rightarrow} \mathcal{S}_k \subset \bigcap_{\substack{\text{Card}(\mathcal{Q}^* \cap \mathcal{Q}) = g \\ \text{Card}(\mathcal{Q} \setminus \mathcal{Q}^*) = b}} \mathcal{S}^{P_q, \mathcal{Q}}$

We address three main questions:

Can one easily characterize $\mathcal{S}^{\text{Algo}}$, $\mathcal{S}_*^{\text{Algo}}$?

The conditions in the standard setup can be generalized and weakened if the support estimate is reliable.

How tight are the conditions defining $\mathcal{S}_*^{\text{Algo}}$?

The coherence condition is tight for all the algorithms. The RIC condition is quasi-tight for Oxx.

Implication between the success of different algorithms ?

Nesting properties are preserved for lp problems and OMP. The behavior of OLS varies in the informed setup.

Further readings...

C. Soussen, R. Gribonval, J. Idier, C. Herzet, «*Joint k-step Analysis of Orthogonal Matching Pursuit and Orthogonal Least Squares*», IEEE Trans. Inf. Theory, May 2013.

C. Herzet , C. Soussen, J. Idier, R. Gribonval, «*Exact Recovery Conditions for Sparse Representations with Partial Support Information*», ArXiV 1305.4008