

Sparsity and Compressed Sensing

Jalal Fadili

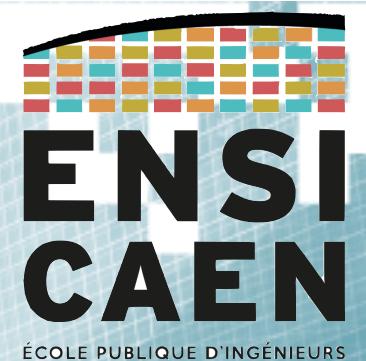
Normandie Université-ENSICAEN, GREYC

<http://www.greyc.ensicaen.fr/~jfadili>

Mathematical coffees 2017



Normandie Université



ÉCOLE PUBLIQUE D'INGÉNIEURS
CENTRE DE RECHERCHE

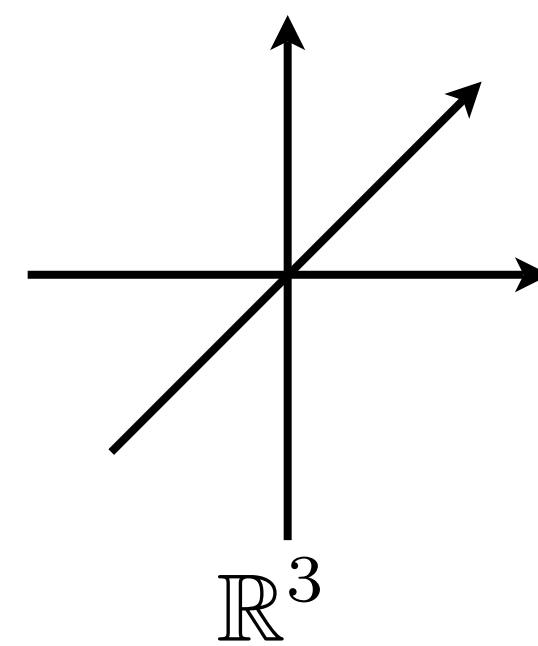
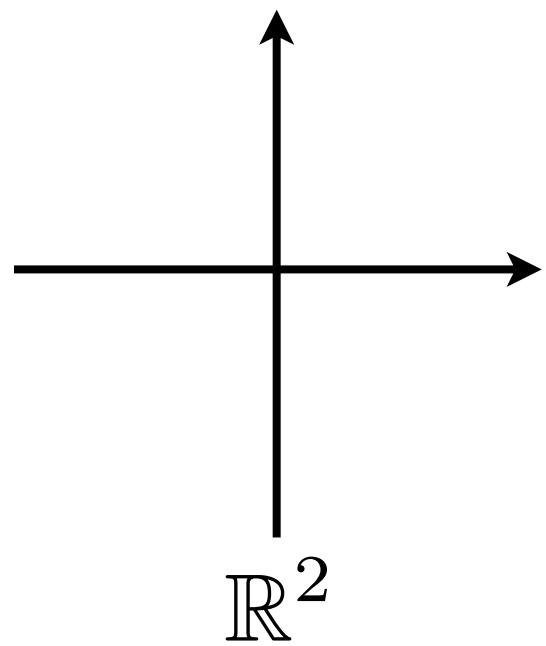
Recap: linear inverse problems

$$y \in \mathbb{R}^m = A x \in \mathbb{R}^n$$

The diagram illustrates a linear transformation. On the left, a vertical vector $y \in \mathbb{R}^m$ is shown as a column of blue squares. In the center, an equals sign ($=$) is positioned above a large, sparse matrix A . The matrix A is composed of small colored squares (blue, red, green, yellow) distributed across its rows and columns. On the right, another vertical vector $x \in \mathbb{R}^n$ is shown as a column of blue squares.

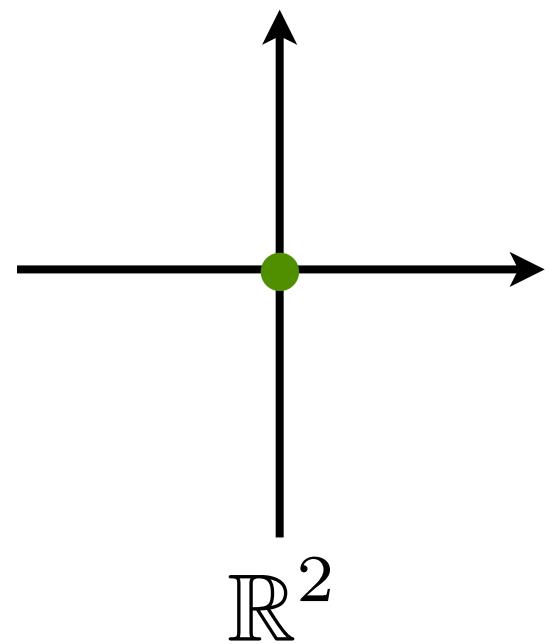
- A is fat (underdetermined system).
 - Solution is not unique (fundamental theorem of linear algebra).
 - Are we stuck ?
 - No if the dimension of x is **intrinsically small**.

Strong notion of sparsity

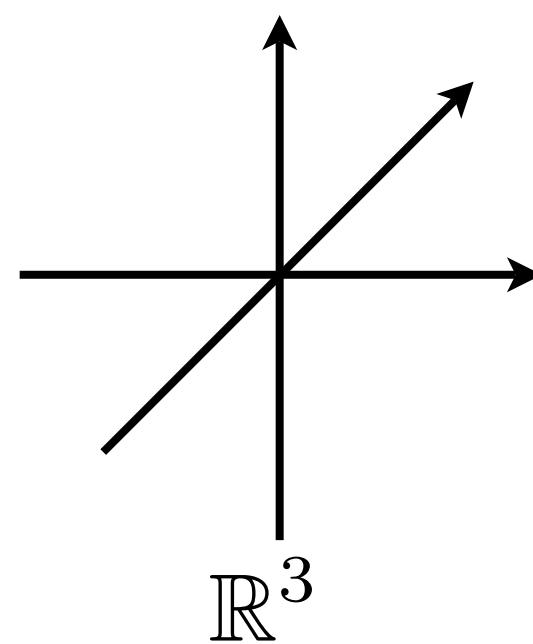


Strong notion of sparsity

- 0-sparse



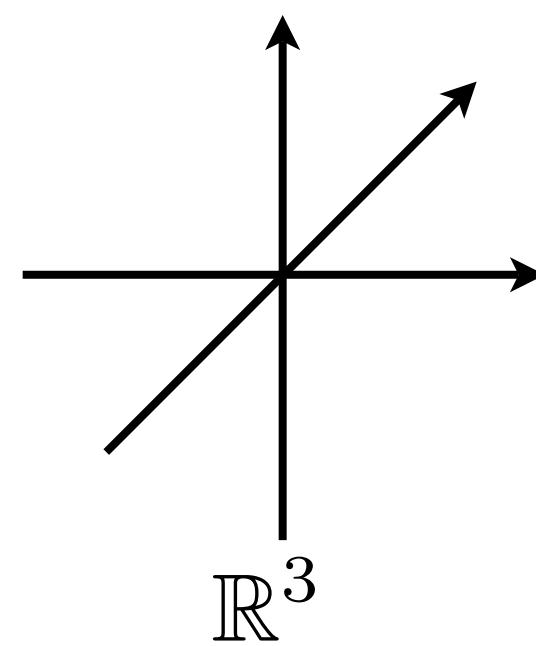
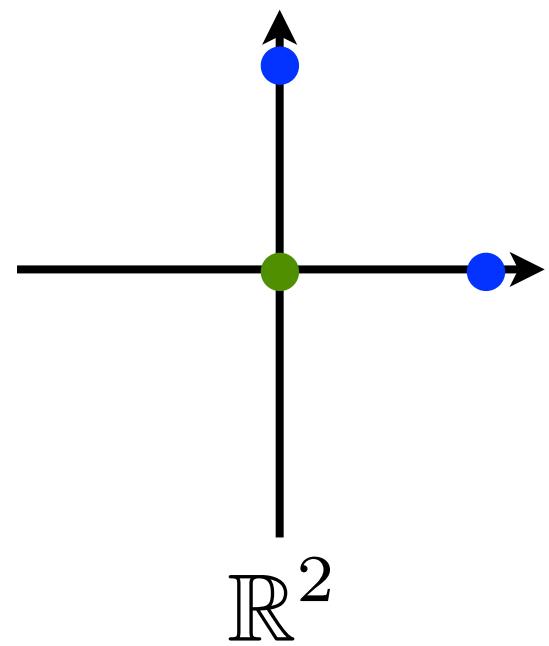
\mathbb{R}^2



\mathbb{R}^3

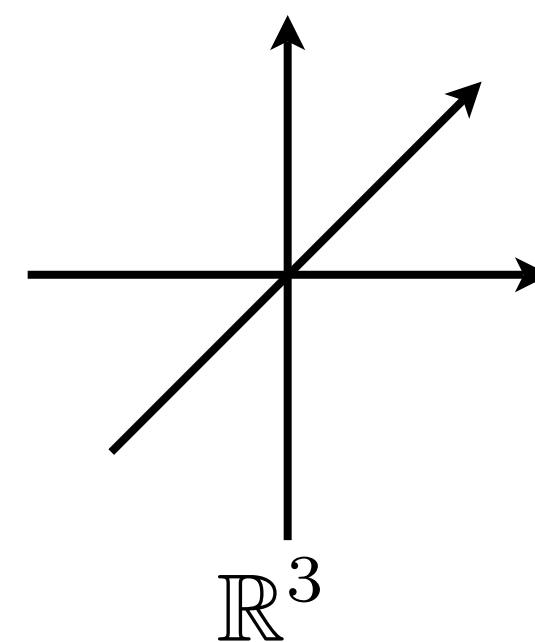
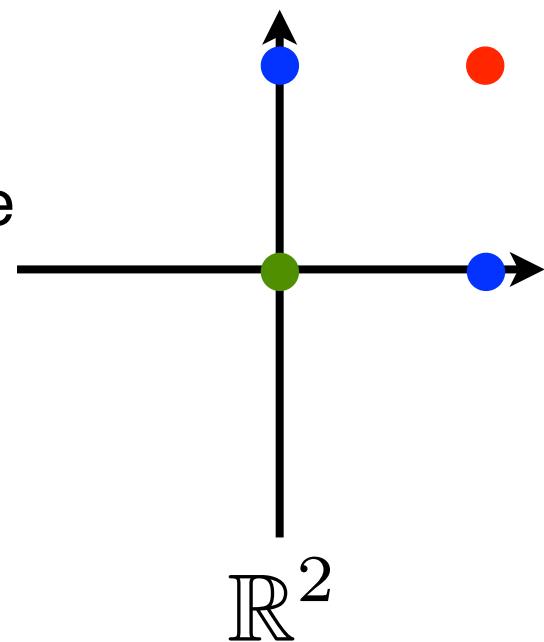
Strong notion of sparsity

- 0-sparse
- 1-sparse



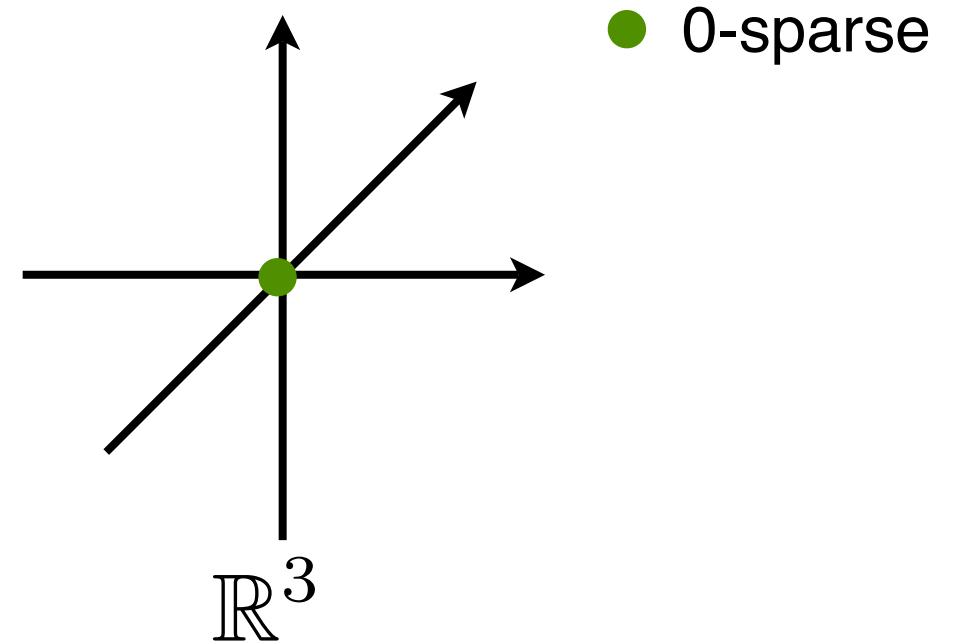
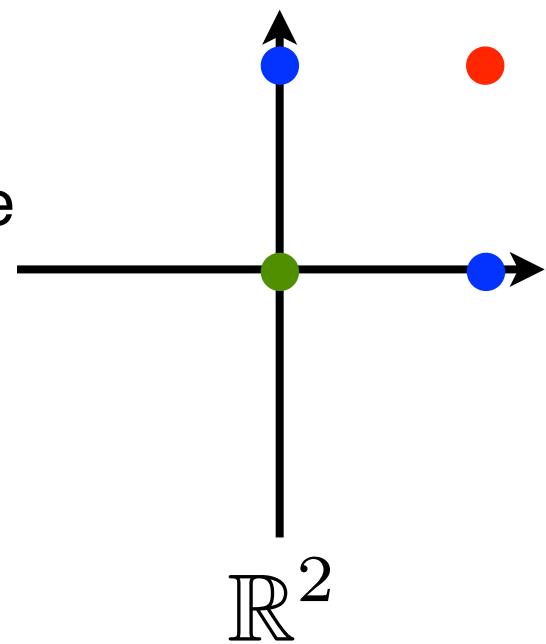
Strong notion of sparsity

- 0-sparse
- 1-sparse
- 2-sparse = dense



Strong notion of sparsity

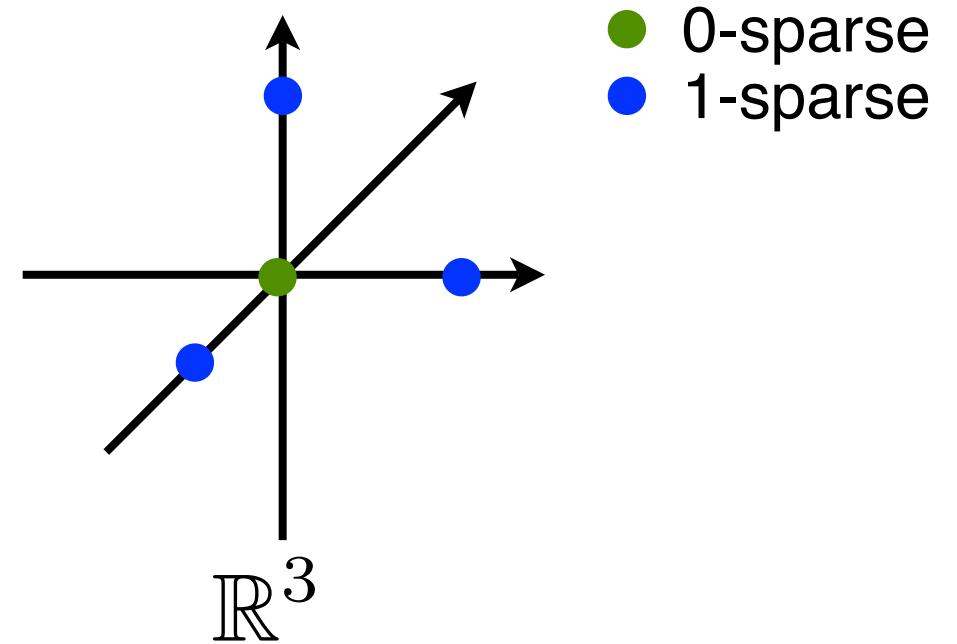
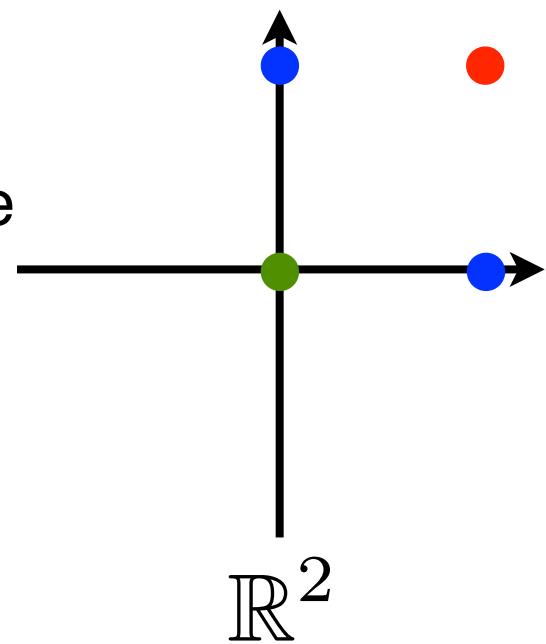
- 0-sparse
- 1-sparse
- 2-sparse = dense



- 0-sparse

Strong notion of sparsity

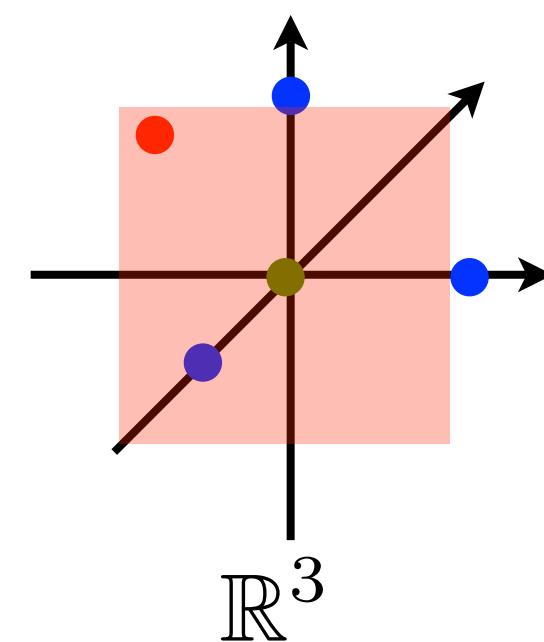
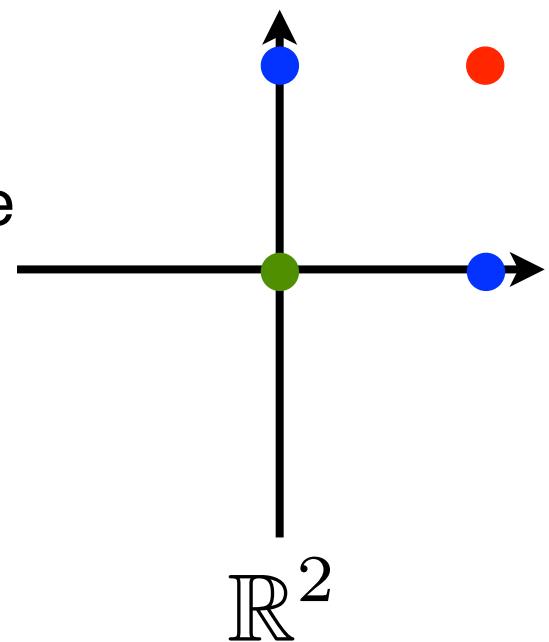
- 0-sparse
- 1-sparse
- 2-sparse = dense



- 0-sparse
- 1-sparse

Strong notion of sparsity

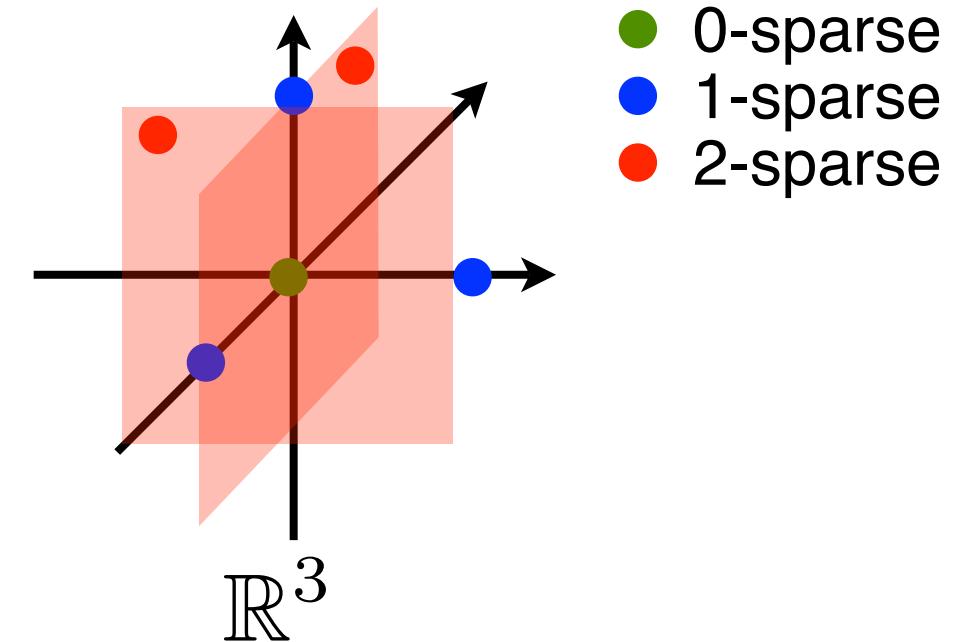
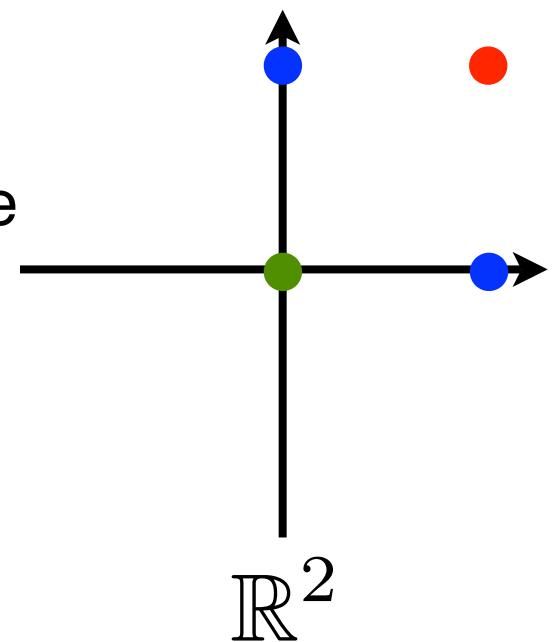
- 0-sparse
 - 1-sparse
 - 2-sparse = dense



- 0-sparse
 - 1-sparse
 - 2-sparse

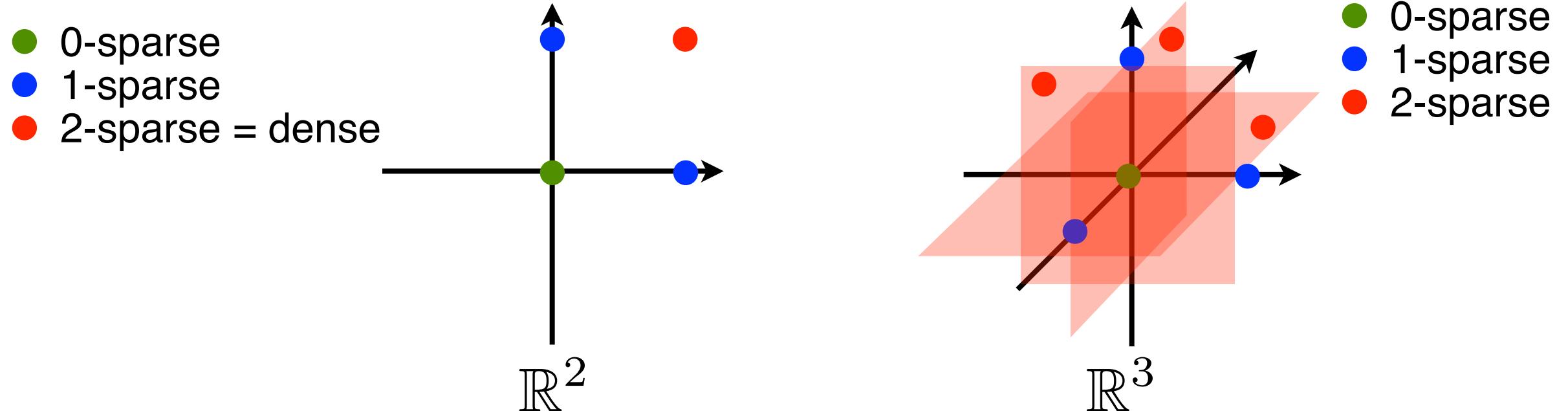
Strong notion of sparsity

- 0-sparse
- 1-sparse
- 2-sparse = dense



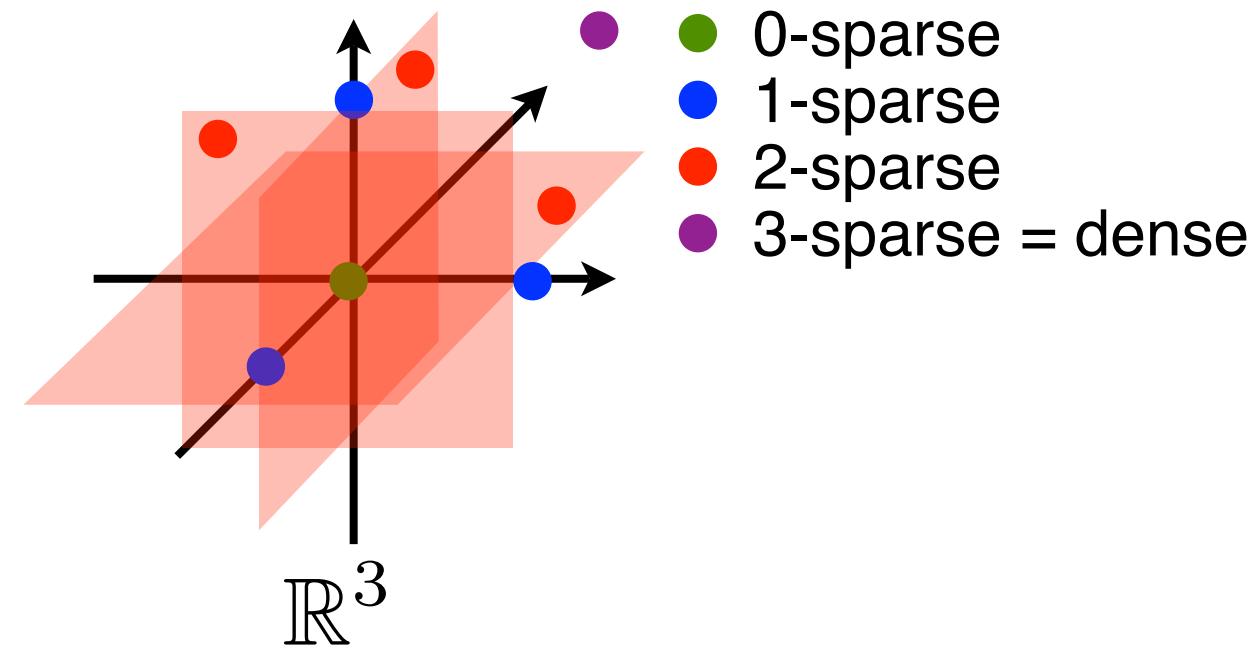
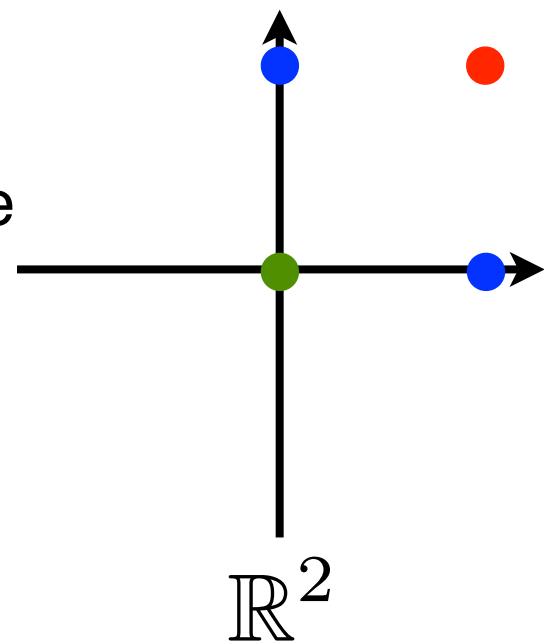
- 0-sparse
- 1-sparse
- 2-sparse

Strong notion of sparsity



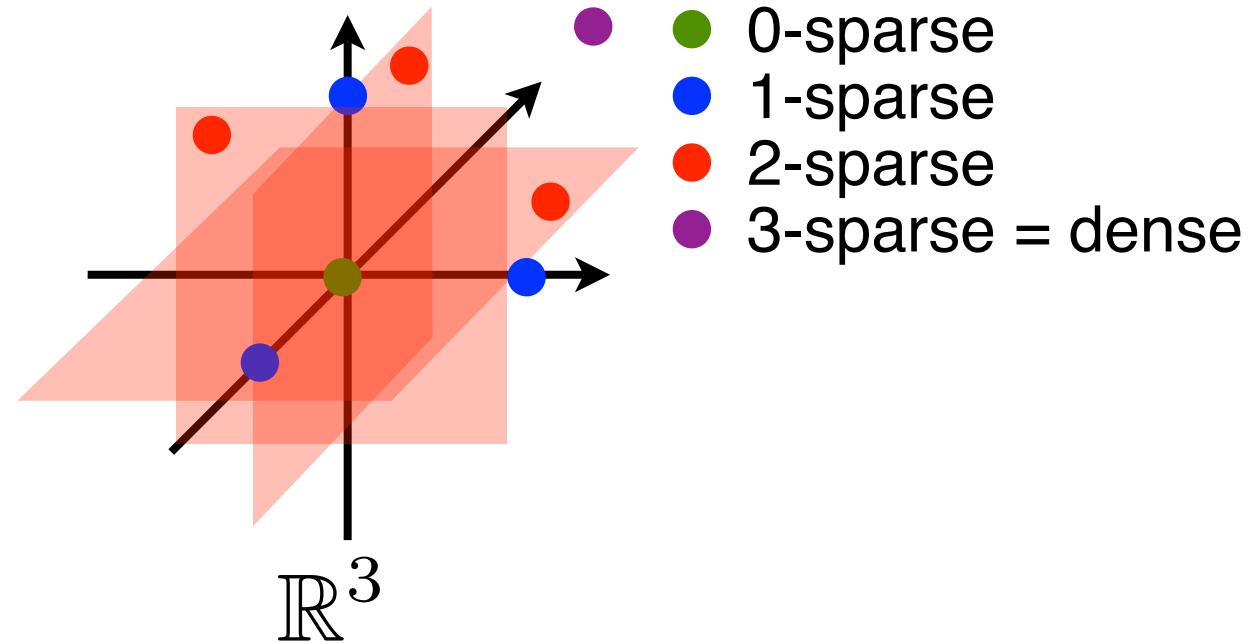
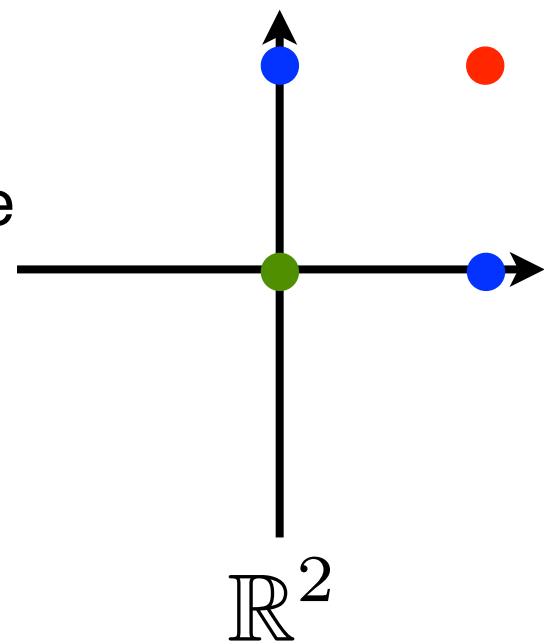
Strong notion of sparsity

- 0-sparse
- 1-sparse
- 2-sparse = dense



Strong notion of sparsity

- 0-sparse
- 1-sparse
- 2-sparse = dense

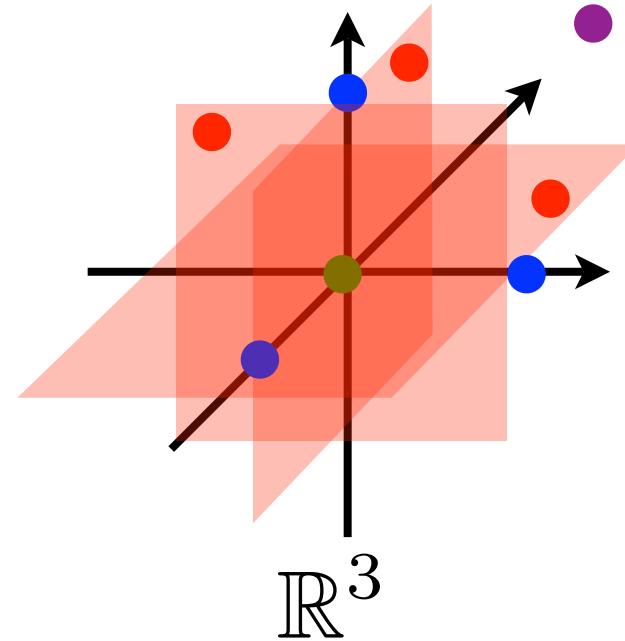
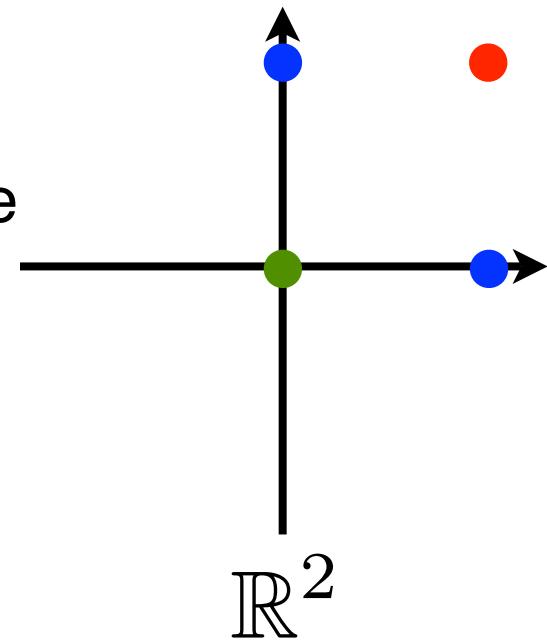


$$\begin{aligned}\text{supp}(x) &= \{i = 1, \dots, n : x_i \neq 0\} \\ \|x\|_0 &= \#\text{supp}(x)\end{aligned}$$

(Not a norm : not positively homogenous)

Strong notion of sparsity

- 0-sparse
- 1-sparse
- 2-sparse = dense



- 0-sparse
- 1-sparse
- 2-sparse
- 3-sparse = dense

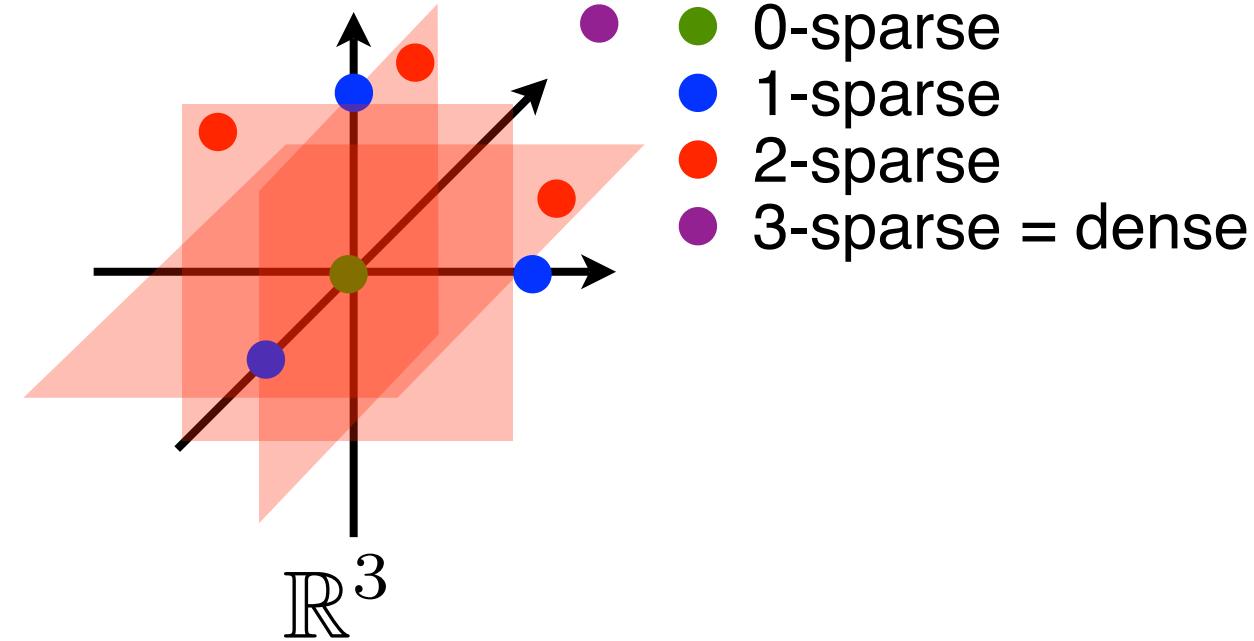
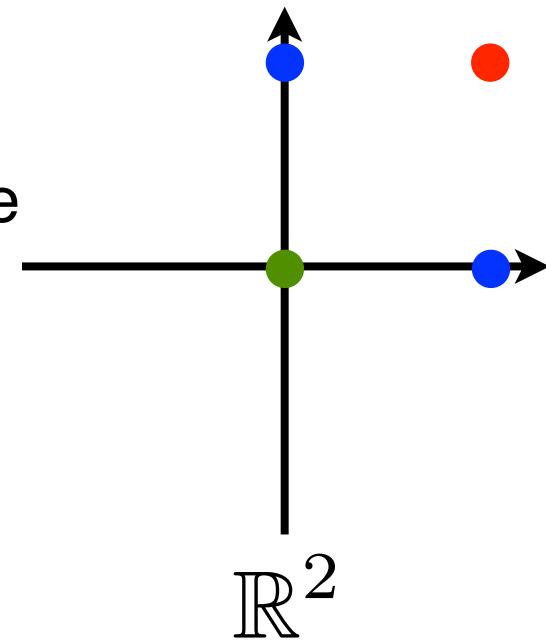
$$\begin{aligned}\text{supp}(x) &= \{i = 1, \dots, n : x_i \neq 0\} \\ \|x\|_0 &= \#\text{supp}(x)\end{aligned}$$

(Not a norm : not positively homogenous)

Definition (Informal) $x \in \mathbb{R}^n$ is sparse iff $\|x\|_0 \ll n$.

Strong notion of sparsity

- 0-sparse
- 1-sparse
- 2-sparse = dense



$$\begin{aligned}\text{supp}(x) &= \{i = 1, \dots, n : x_i \neq 0\} \\ \|x\|_0 &= \#\text{supp}(x)\end{aligned}$$

(Not a norm : not positively homogenous)

Definition (Informal) $x \in \mathbb{R}^n$ is sparse iff $\|x\|_0 \ll n$.

Model of s -sparse vectors : a union of subspaces

$$\Sigma_s = \bigcup_i \{V_i = \text{span}((e_j)_{1 \leq j \leq n}) : \dim(V_i) = s\}.$$

Weak notion of sparsity

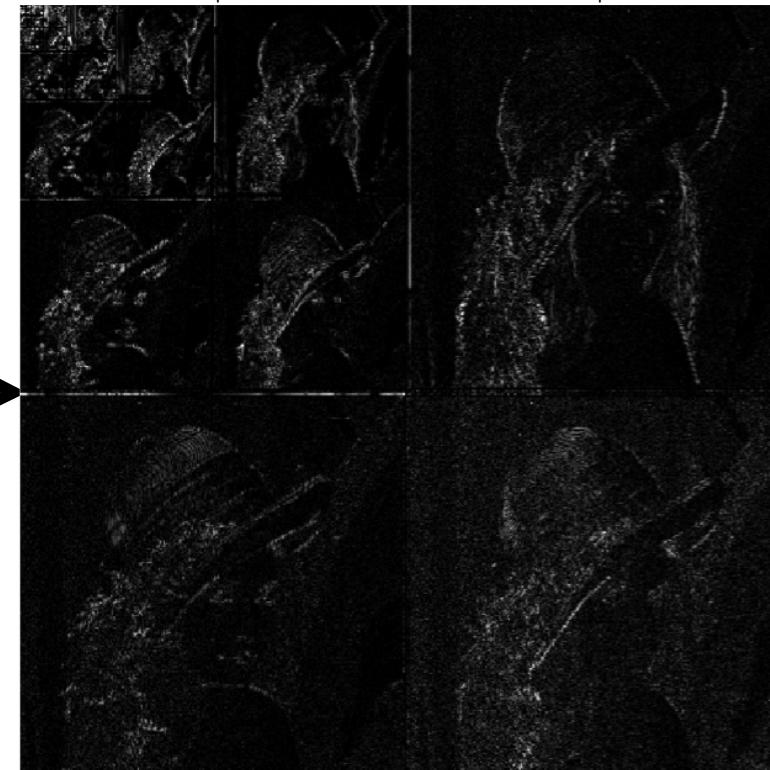
- In nature, signals, images, information, are not (strongly) sparse .

Weak notion of sparsity

- In nature, signals, images, information, are not (strongly) sparse .



Wavelet transform



$\gg 0$ white

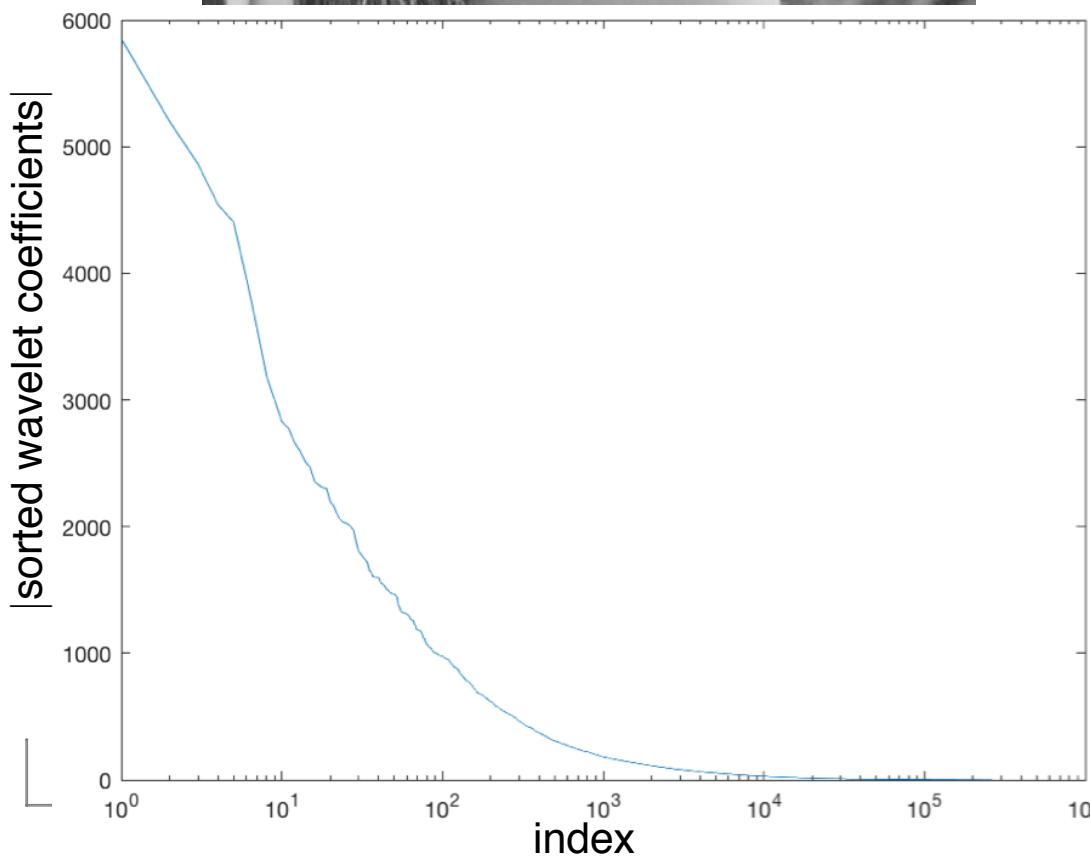
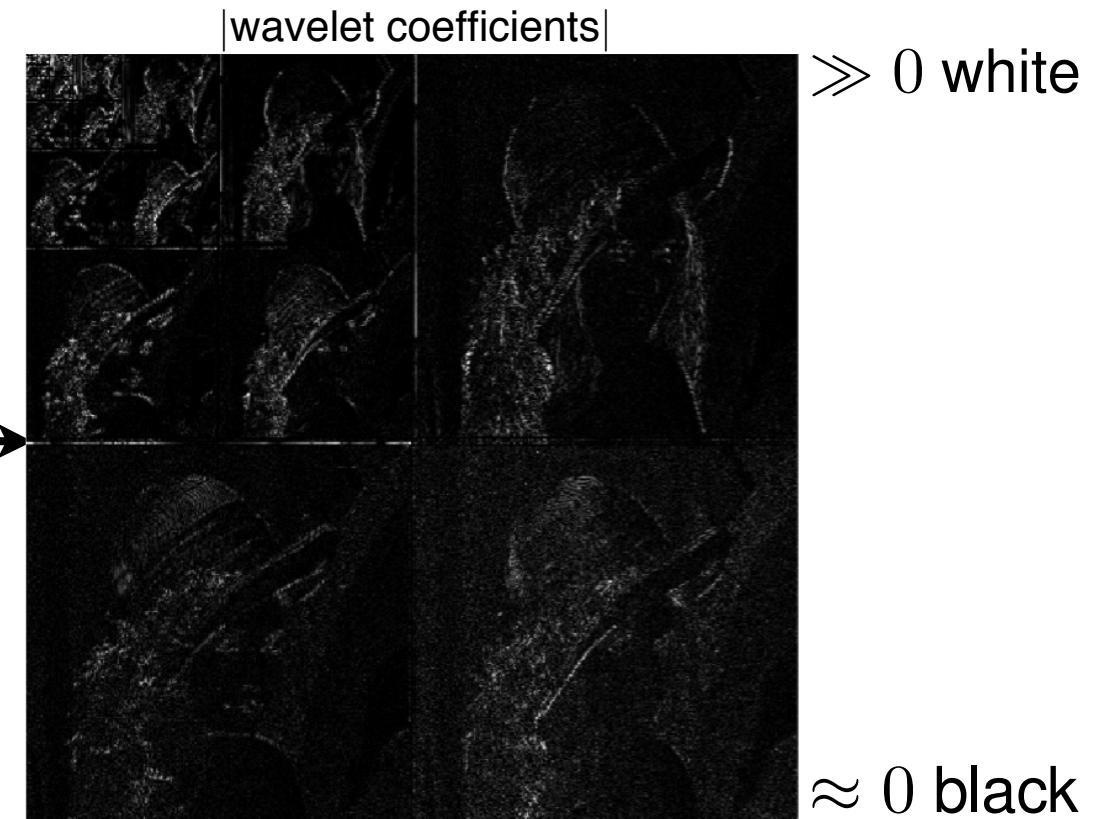
≈ 0 black

Weak notion of sparsity

- In nature, signals, images, information, are not (strongly) sparse .



Wavelet transform

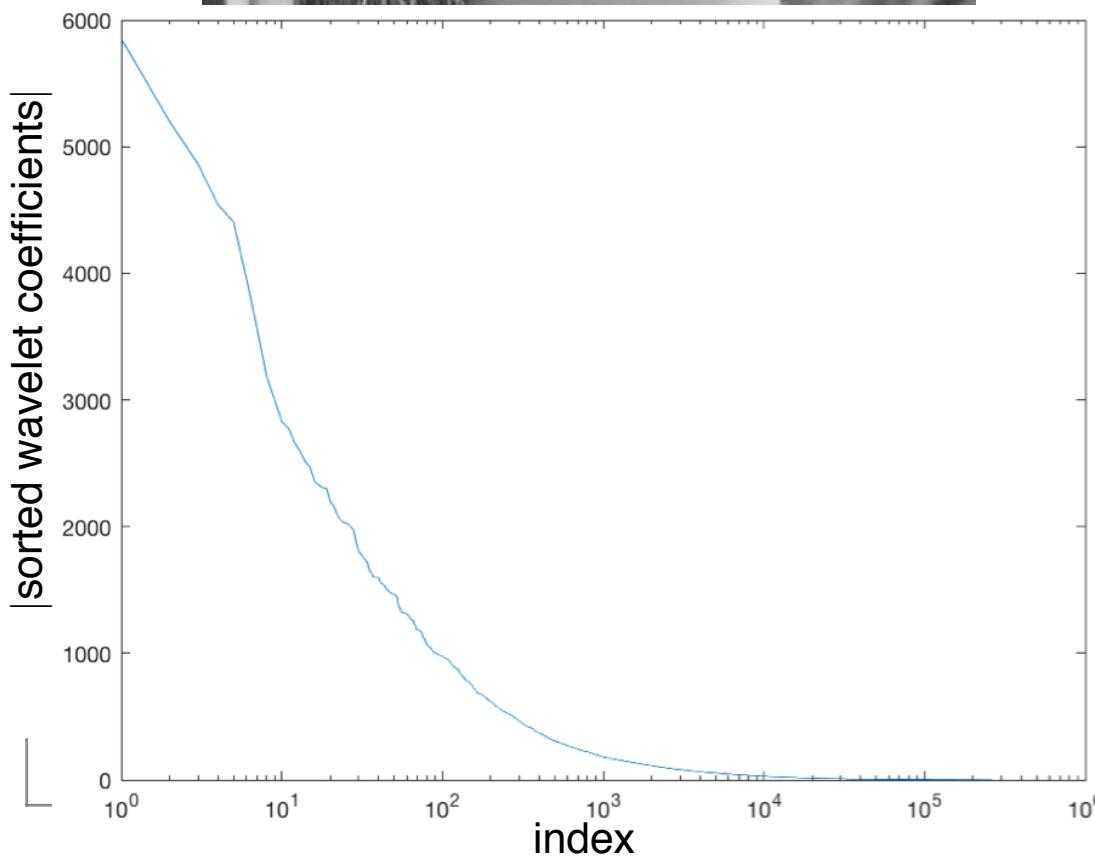
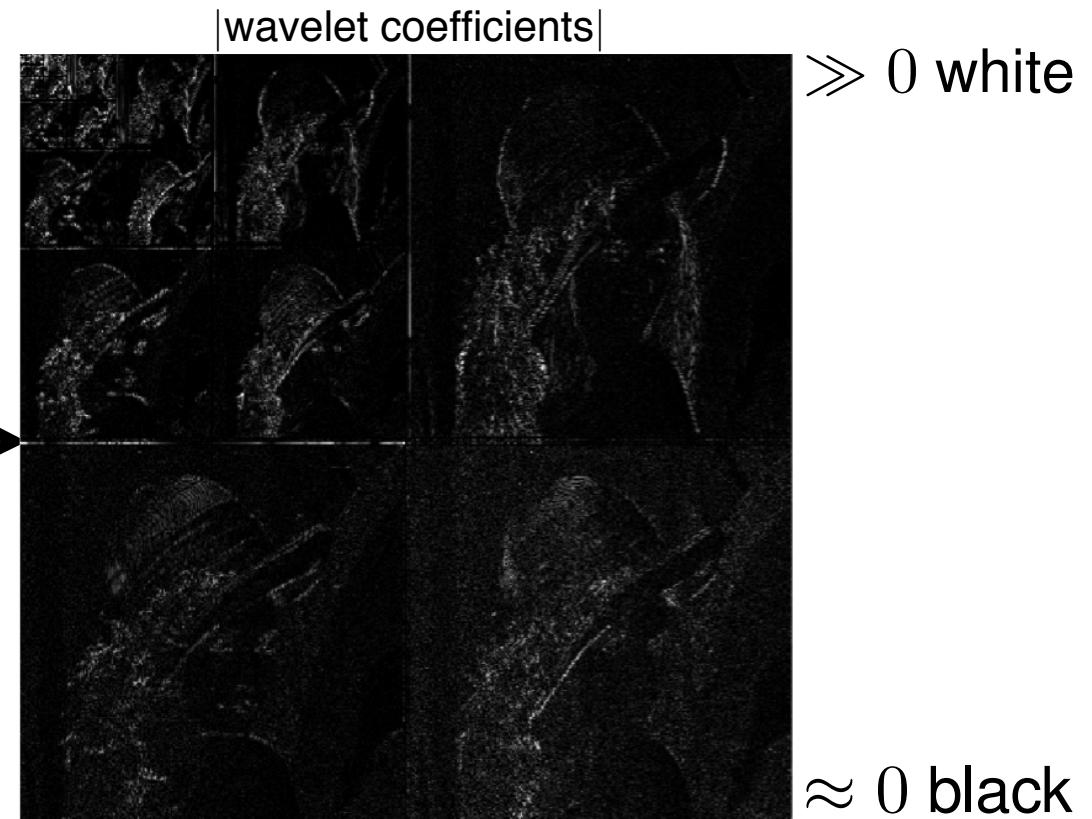


Weak notion of sparsity

- In nature, signals, images, information, are not (strongly) sparse .



Wavelet transform



Definition (Informal) $x \in \mathbb{R}^n$ is weakly sparse iff $|x_{(i)}|$ decreases fast enough with i .

From now on, sparsity is intended in strong sense

What sparsity good for ?

Solve $y = Ax$ where x is sparse

- If $\|x\|_0 \leq m$ and A_I is full-rank ($I \stackrel{\text{def}}{=} \text{supp}(x)$), we are done.
- Indeed, at least as many equations as unknowns :

$$y = A_I x_I.$$

- In practice, the support I is not known.
- We have to infer it from the sole knowledge of y and A .

Regularization

Solve $y = Ax$ where x is sparse

Regularization

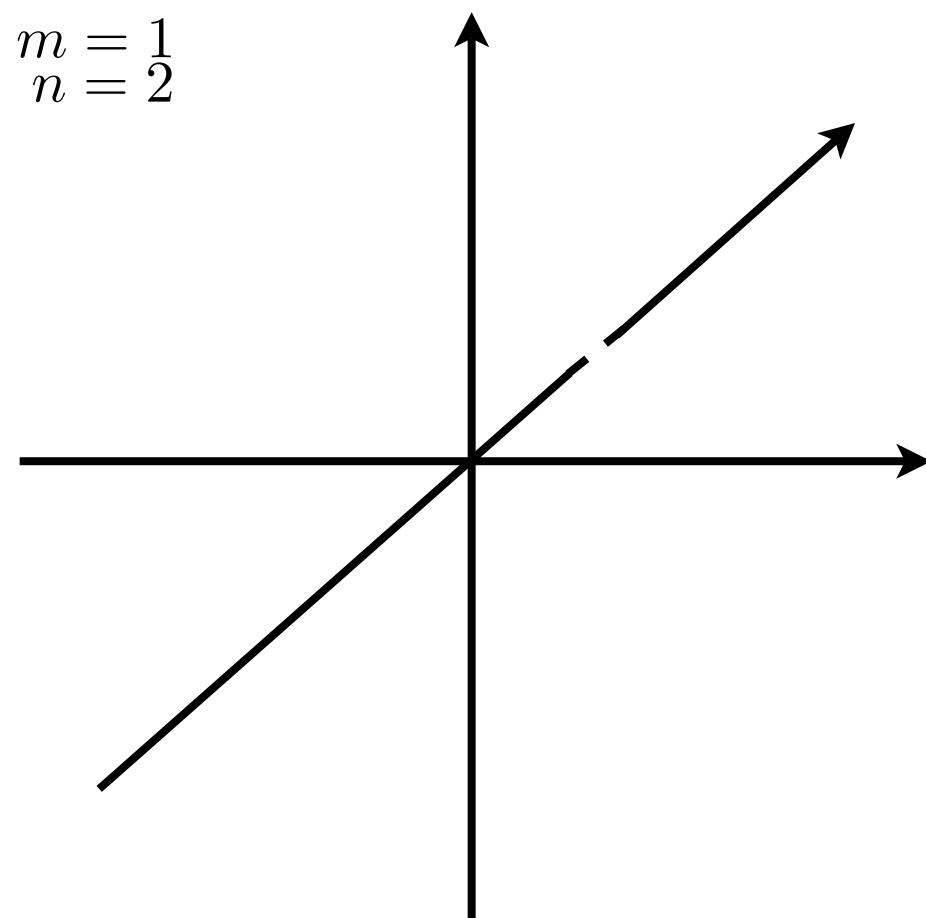
Solve $y = Ax$ where x is sparse

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \text{ such that } y = Ax$$

Regularization

Solve $y = Ax$ where x is sparse

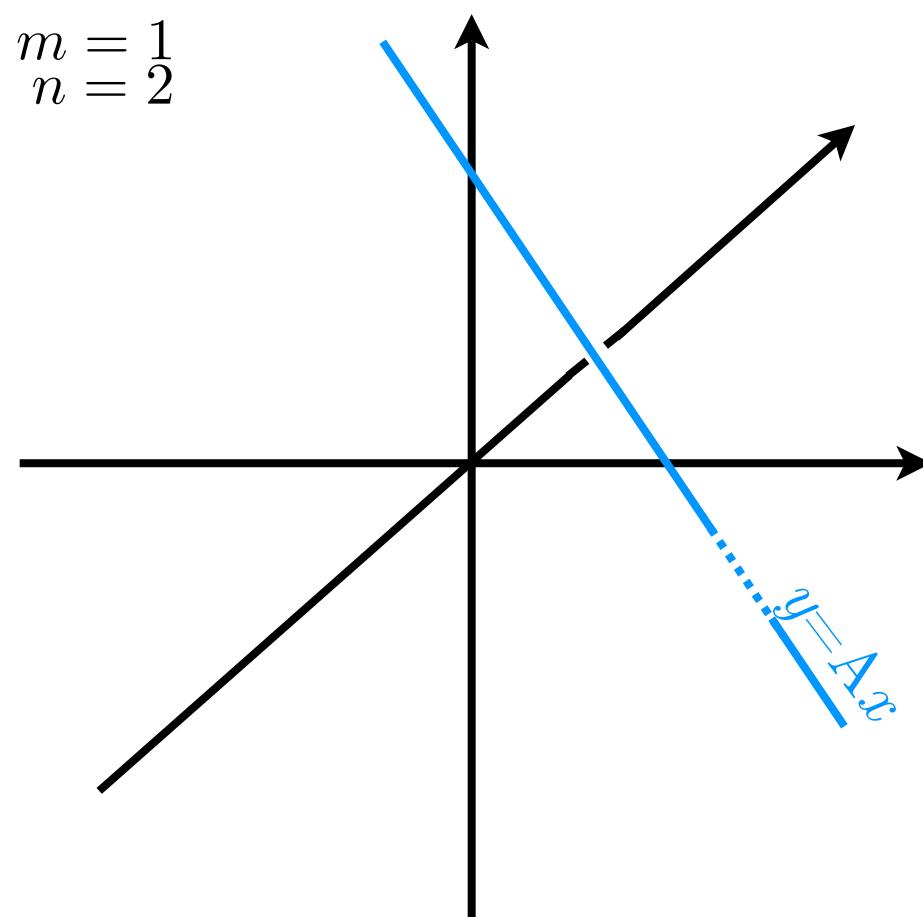
$$\min_{x \in \mathbb{R}^n} \|x\|_0 \text{ such that } y = Ax$$



Regularization

Solve $y = Ax$ where x is sparse

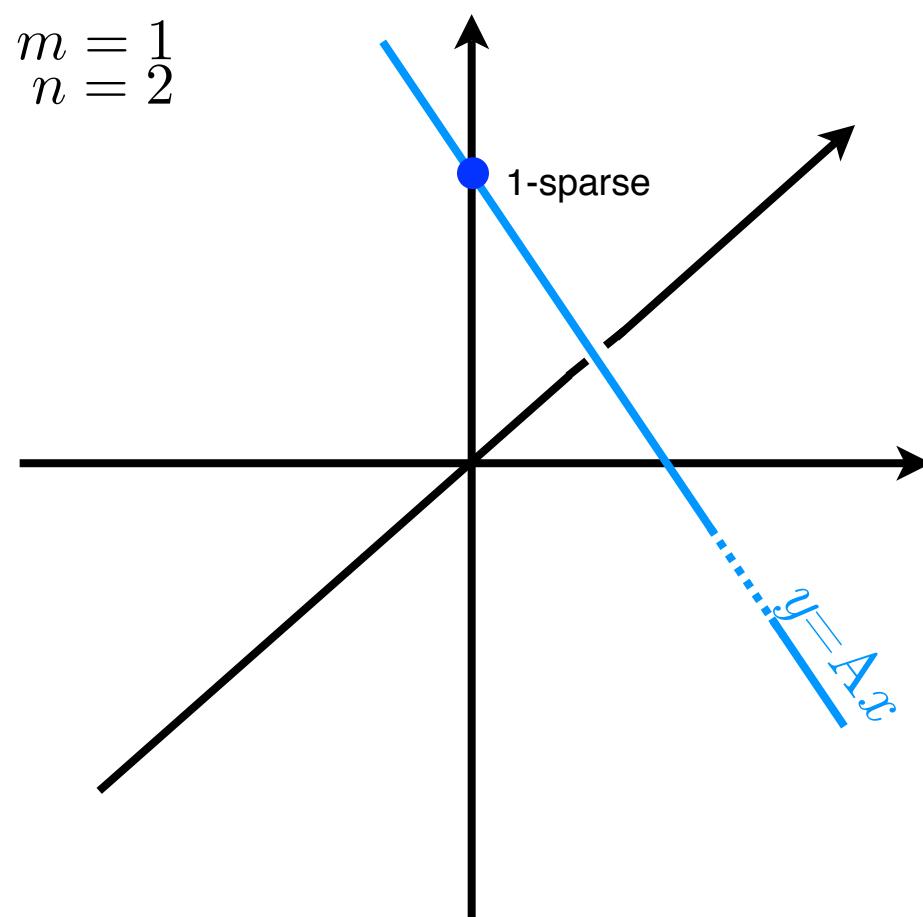
$$\min_{x \in \mathbb{R}^n} \|x\|_0 \text{ such that } y = Ax$$



Regularization

Solve $y = Ax$ where x is sparse

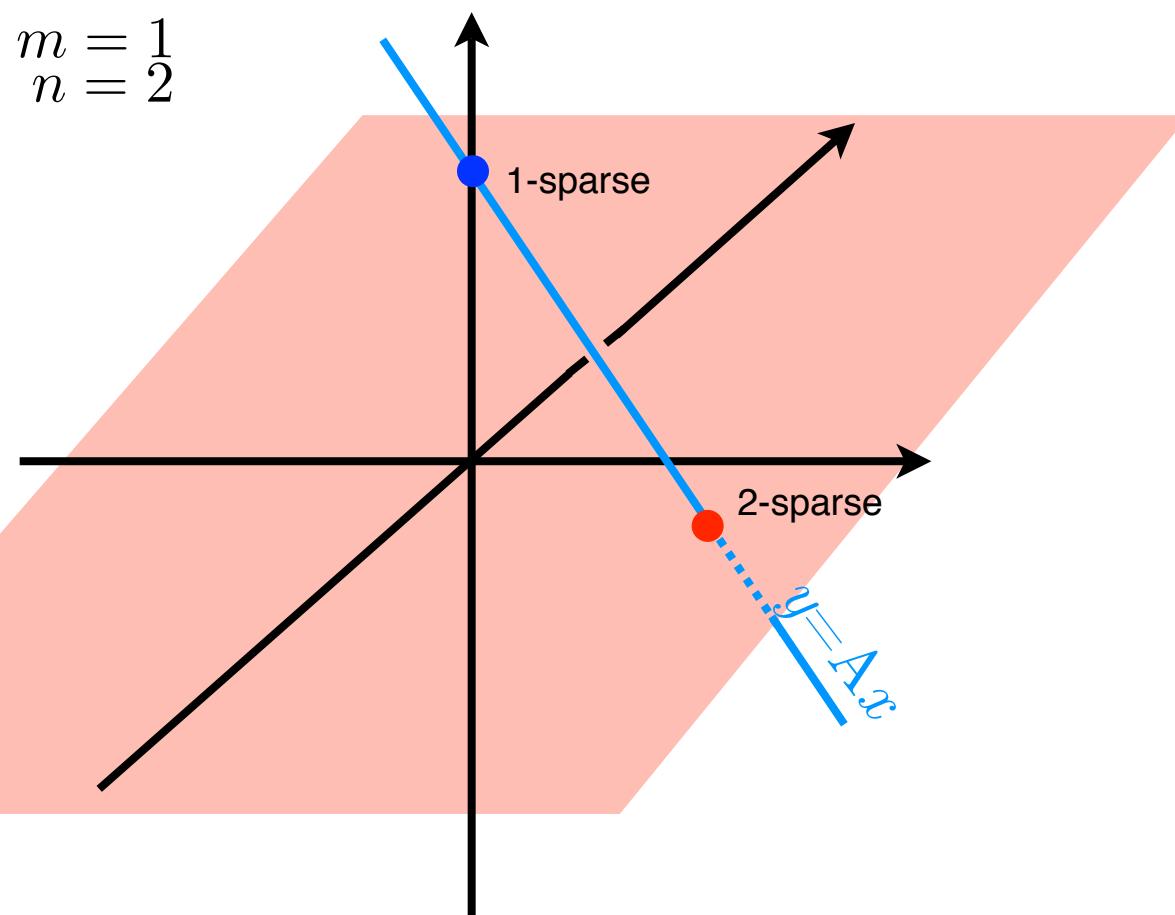
$$\min_{x \in \mathbb{R}^n} \|x\|_0 \text{ such that } y = Ax$$



Regularization

Solve $y = Ax$ where x is sparse

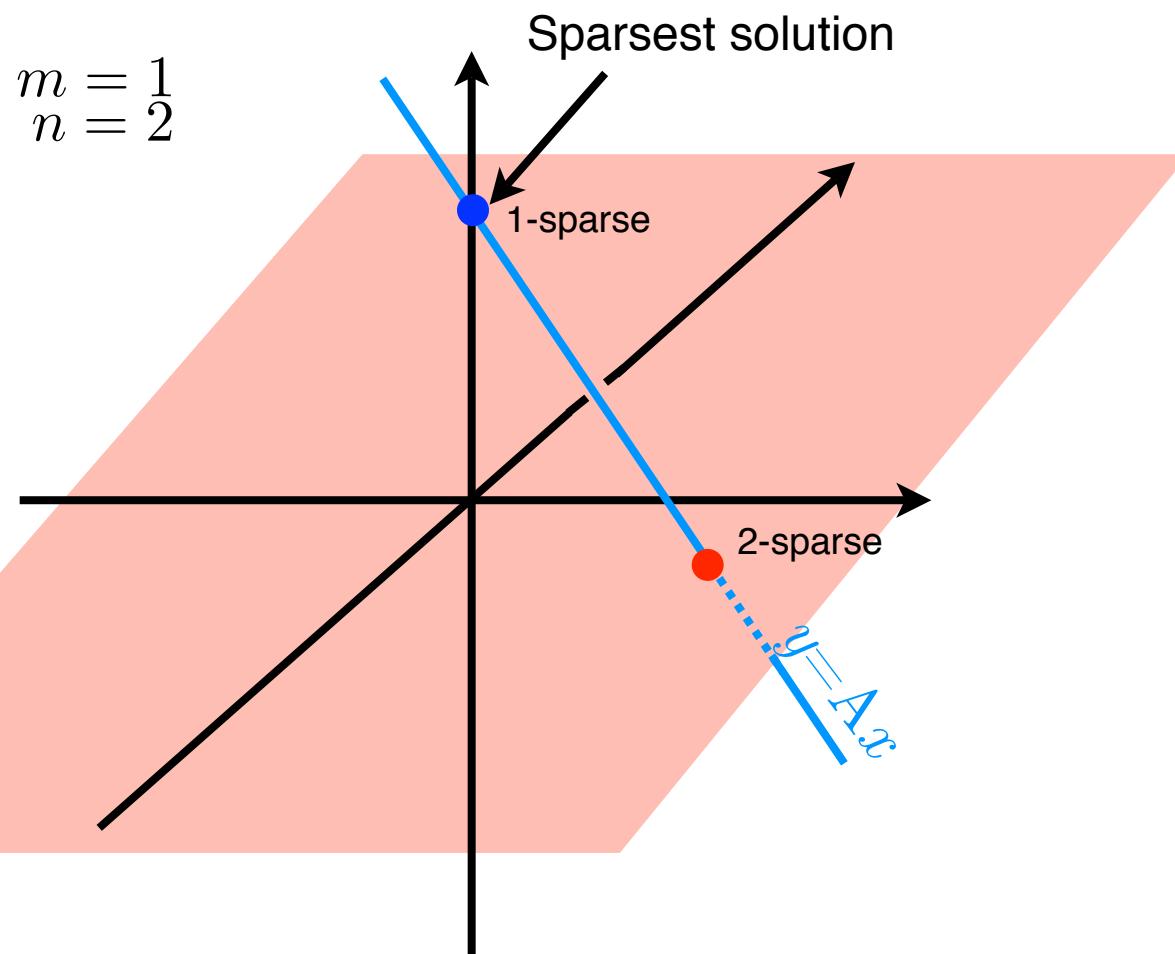
$$\min_{x \in \mathbb{R}^n} \|x\|_0 \text{ such that } y = Ax$$



Regularization

Solve $y = Ax$ where x is sparse

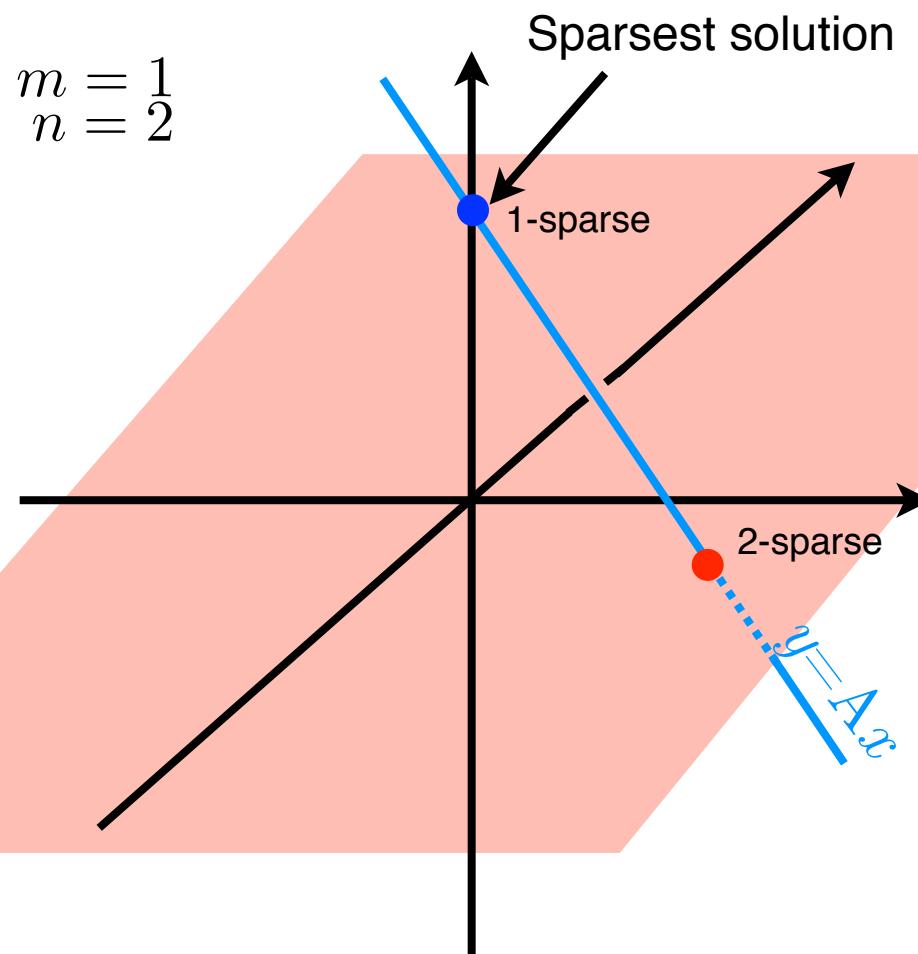
$$\min_{x \in \mathbb{R}^n} \|x\|_0 \text{ such that } y = Ax$$



Regularization

Solve $y = Ax$ where x is sparse

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \text{ such that } y = Ax$$



- Not convex, not even continuous.
- In fact, this is a combinatorial NP-hard problem.
- Can we find a viable alternative ?

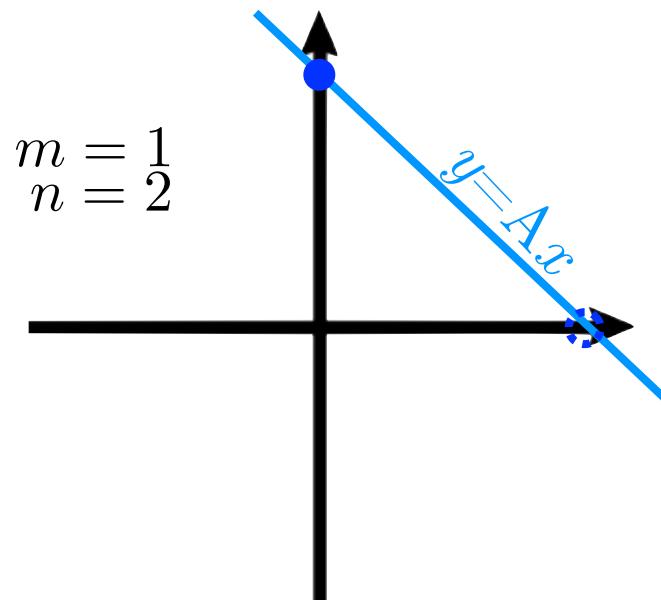
Relaxation

Solve $y = Ax$ where x is sparse

Relaxation

Solve $y = Ax$ where x is sparse

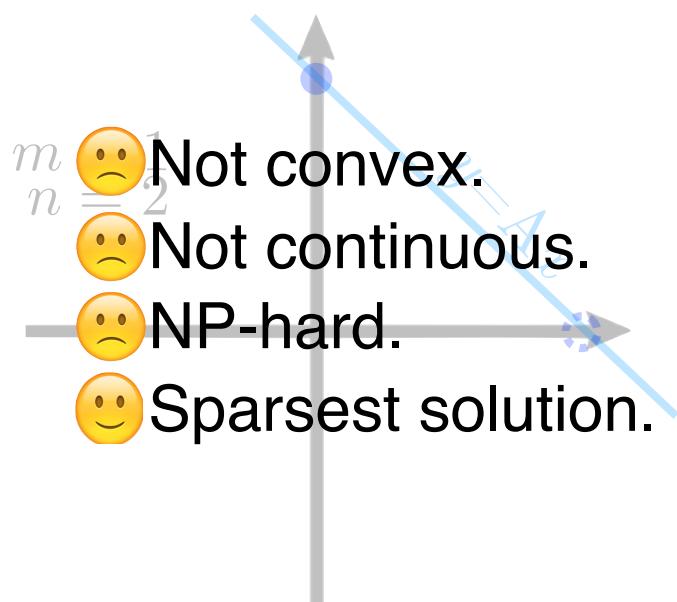
$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{s.t. } y = Ax$$



Relaxation

Solve $y = Ax$ where x is sparse

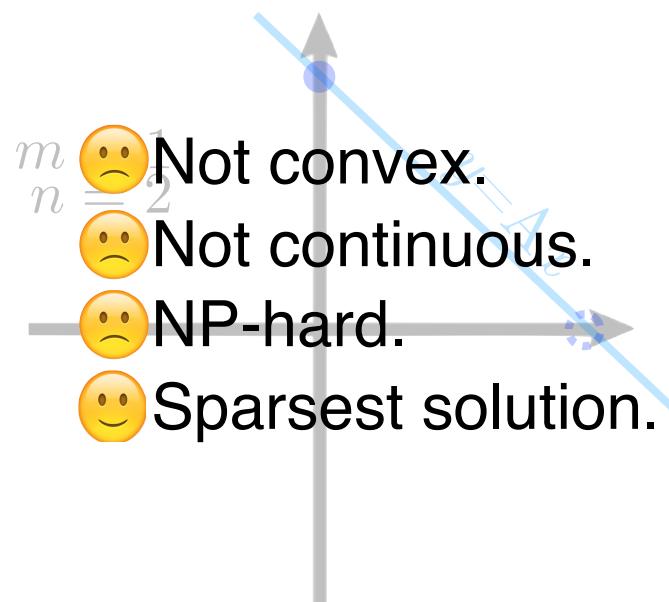
$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{s.t. } y = Ax$$



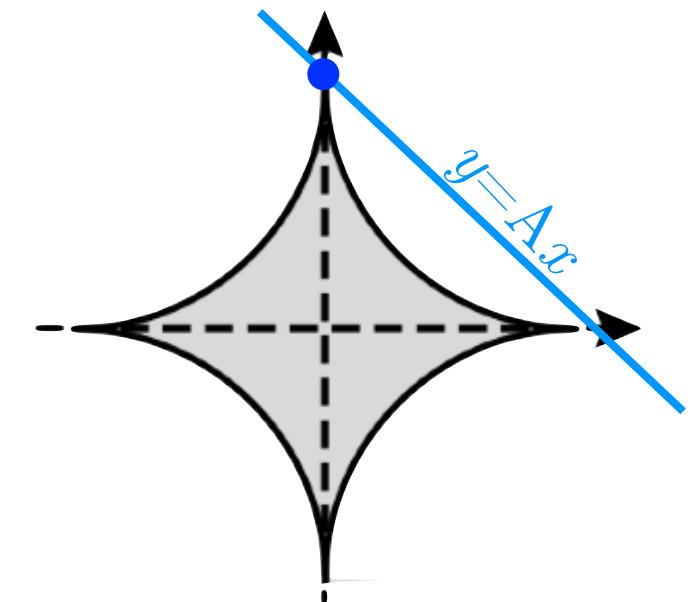
Relaxation

Solve $y = Ax$ where x is sparse

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{s.t. } y = Ax$$



$$\min_{x \in \mathbb{R}^n} \|x\|_{0.5} \quad \text{s.t. } y = Ax$$

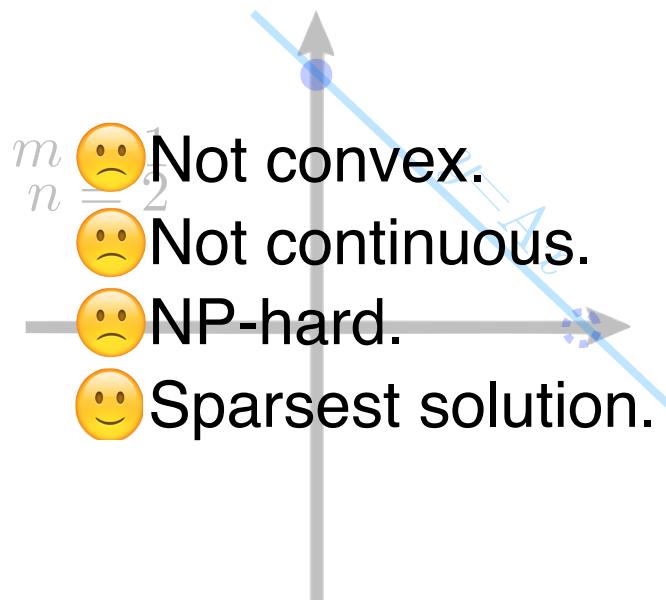


$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$$

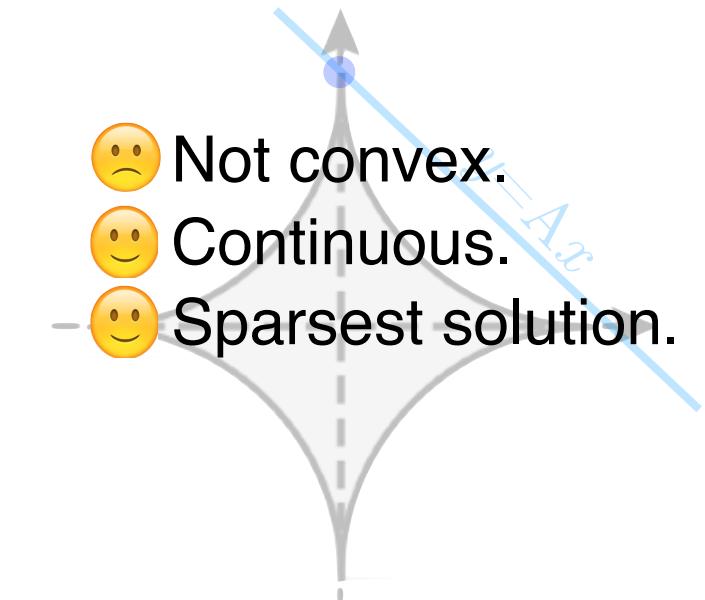
Relaxation

Solve $y = Ax$ where x is sparse

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{s.t. } y = Ax$$



$$\min_{x \in \mathbb{R}^n} \|x\|_{0.5} \quad \text{s.t. } y = Ax$$

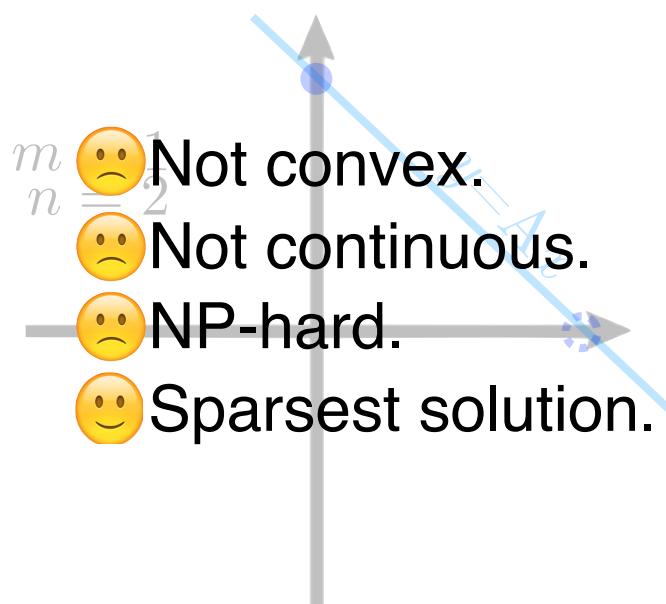


$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$$

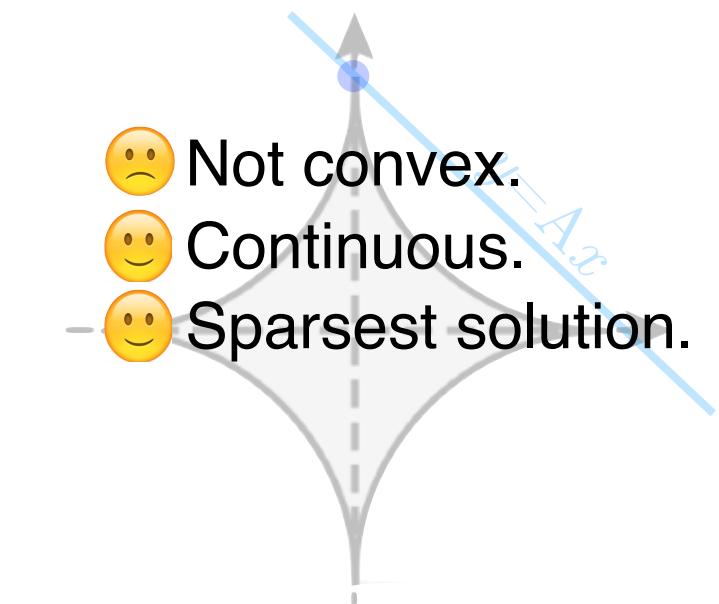
Relaxation

Solve $y = Ax$ where x is sparse

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \text{ s.t. } y = Ax$$

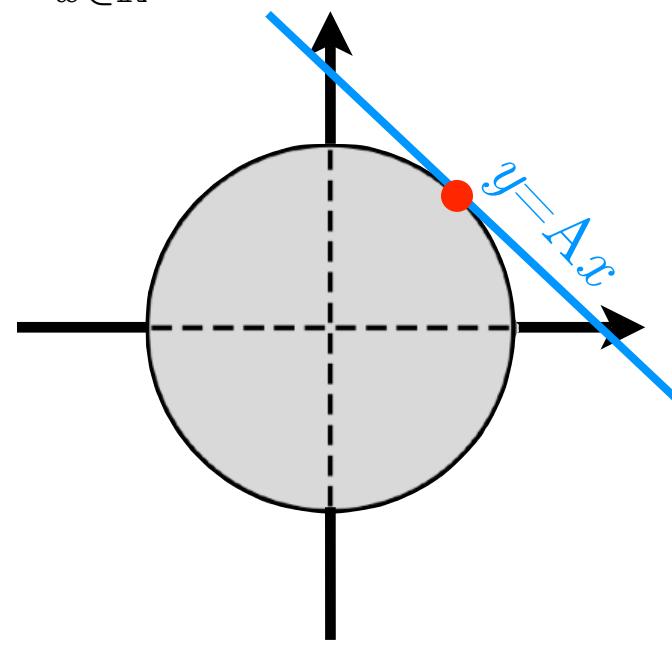


$$\min_{x \in \mathbb{R}^n} \|x\|_{0.5} \text{ s.t. } y = Ax$$



$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$$

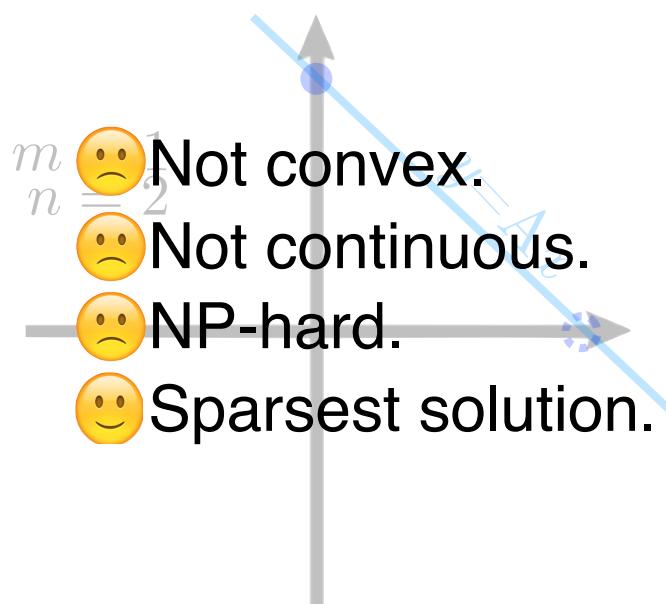
$$\min_{x \in \mathbb{R}^n} \|x\|_2 \text{ s.t. } y = Ax$$



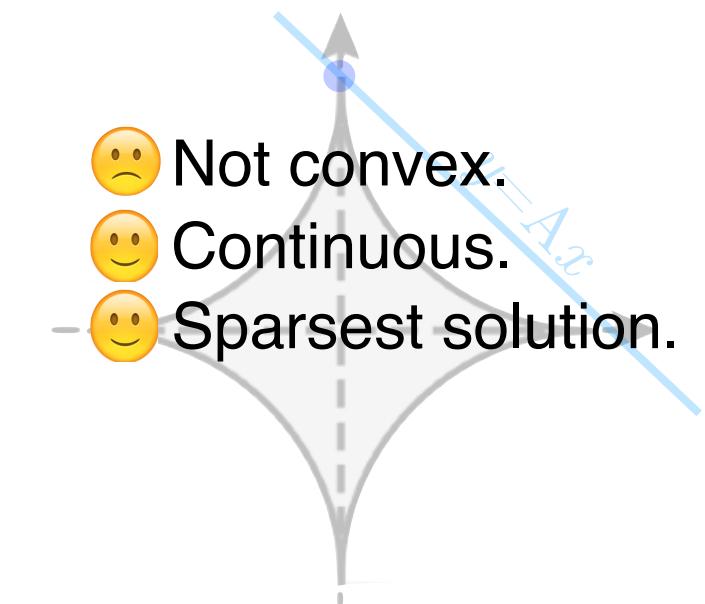
Relaxation

Solve $y = Ax$ where x is sparse

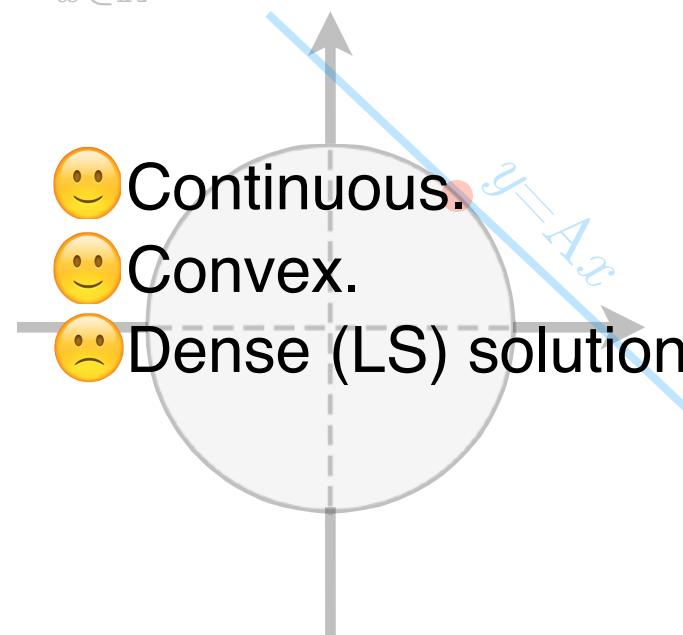
$$\min_{x \in \mathbb{R}^n} \|x\|_0 \text{ s.t. } y = Ax$$



$$\min_{x \in \mathbb{R}^n} \|x\|_{0.5} \text{ s.t. } y = Ax$$



$$\min_{x \in \mathbb{R}^n} \|x\|_2 \text{ s.t. } y = Ax$$

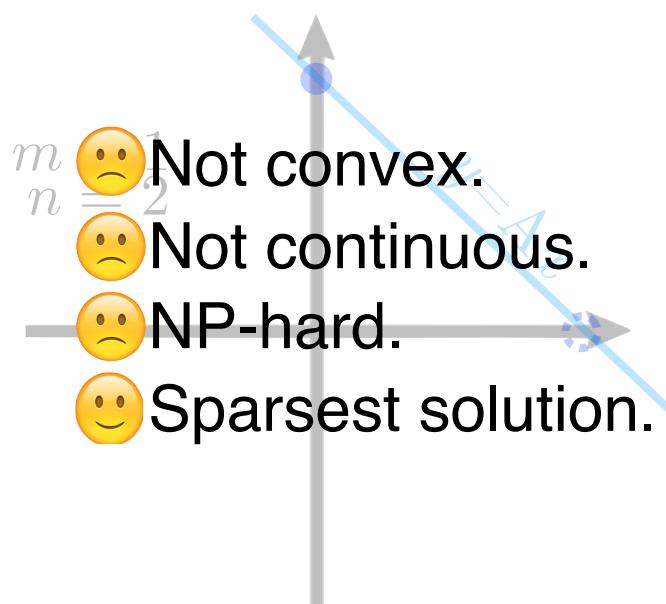


$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$$

Relaxation

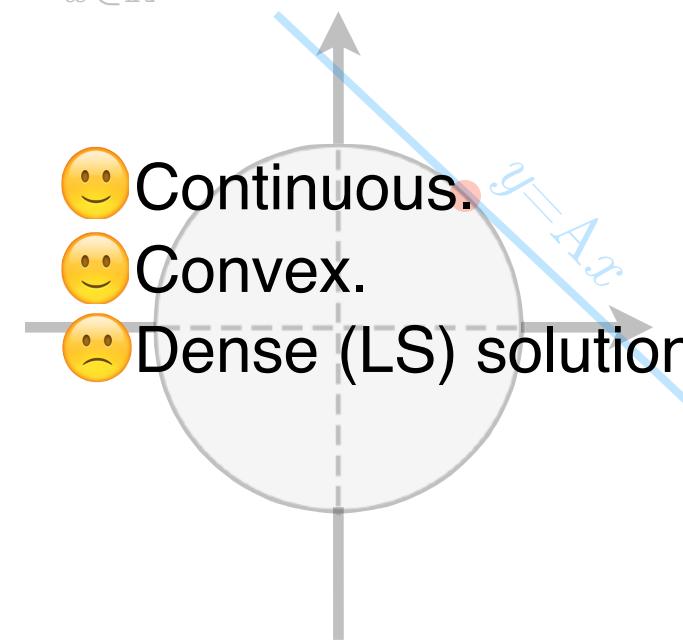
Solve $y = Ax$ where x is sparse

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \text{ s.t. } y = Ax$$

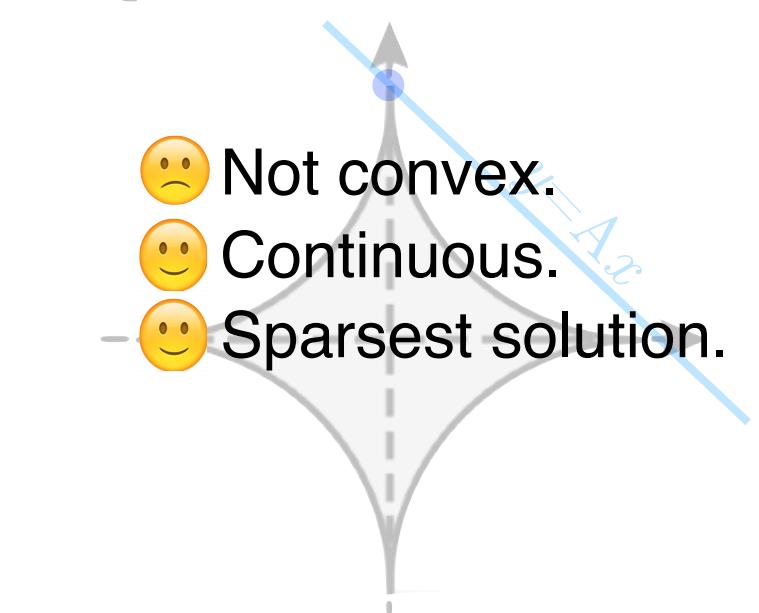


$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$$

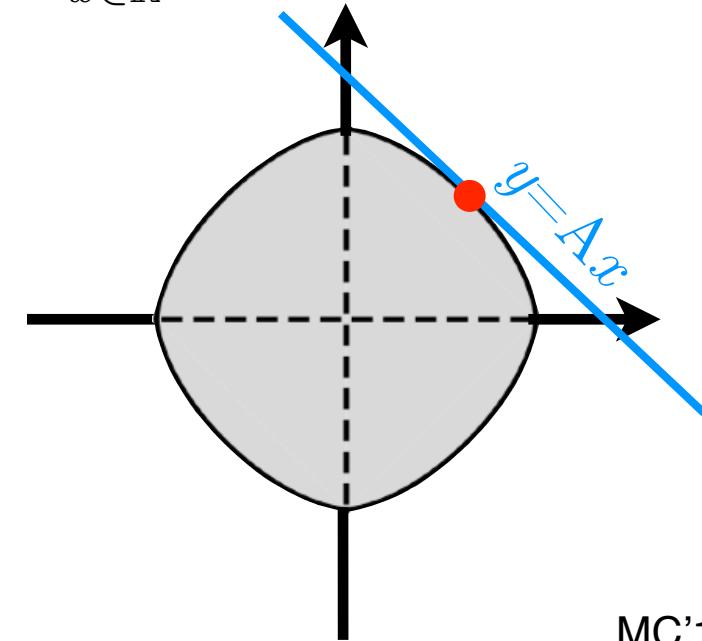
$$\min_{x \in \mathbb{R}^n} \|x\|_2 \text{ s.t. } y = Ax$$



$$\min_{x \in \mathbb{R}^n} \|x\|_{0.5} \text{ s.t. } y = Ax$$



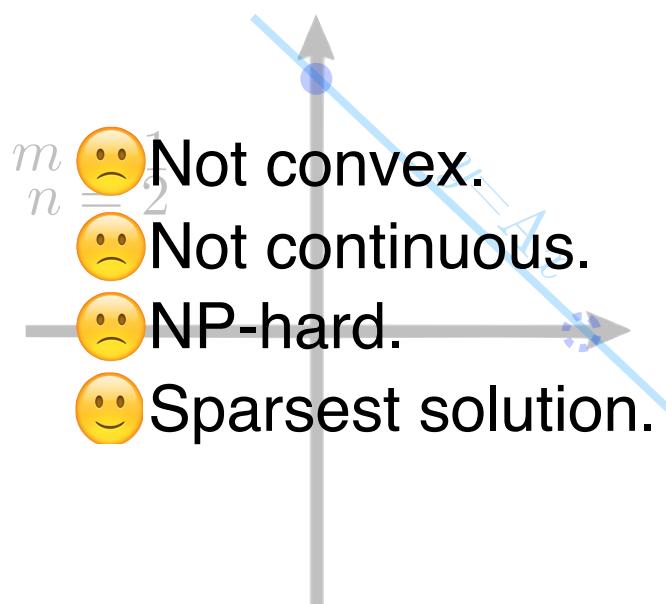
$$\min_{x \in \mathbb{R}^n} \|x\|_{1.5} \text{ s.t. } y = Ax$$



Relaxation

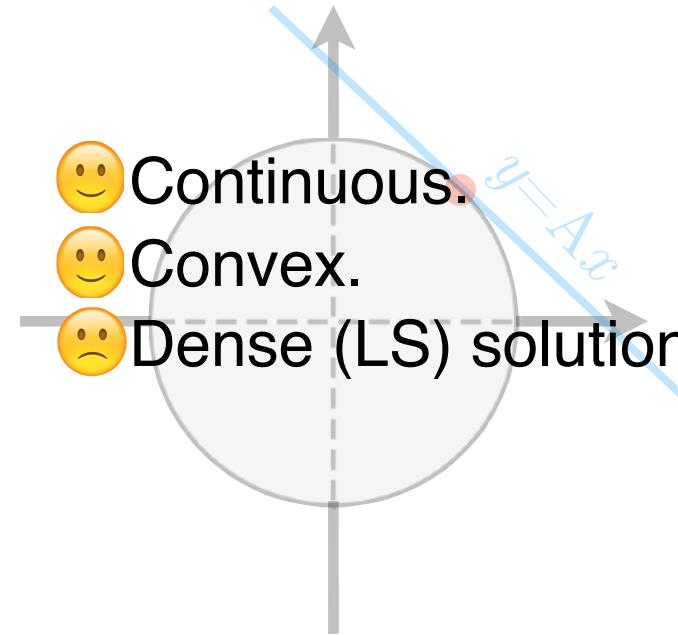
Solve $y = Ax$ where x is sparse

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \text{ s.t. } y = Ax$$

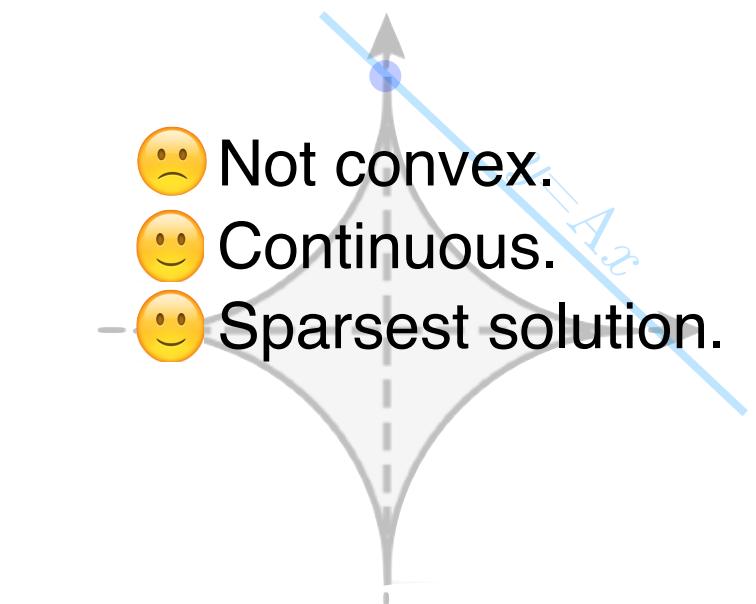


$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$$

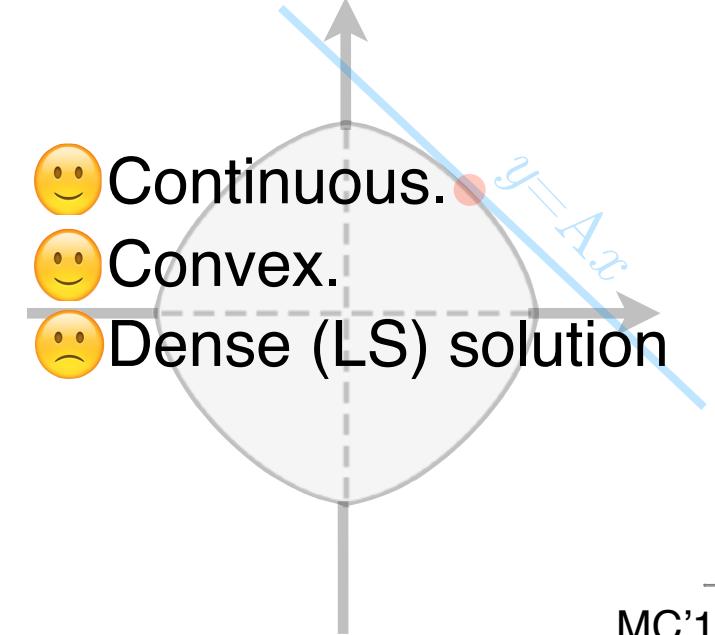
$$\min_{x \in \mathbb{R}^n} \|x\|_2 \text{ s.t. } y = Ax$$



$$\min_{x \in \mathbb{R}^n} \|x\|_{0.5} \text{ s.t. } y = Ax$$



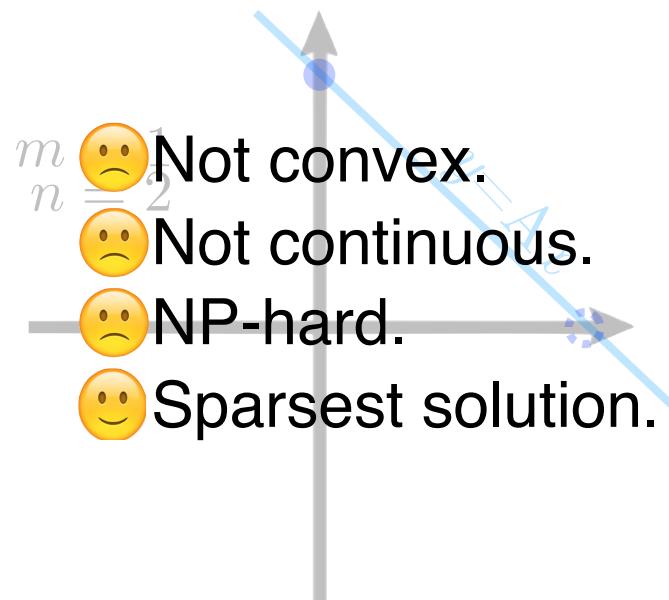
$$\min_{x \in \mathbb{R}^n} \|x\|_{1.5} \text{ s.t. } y = Ax$$



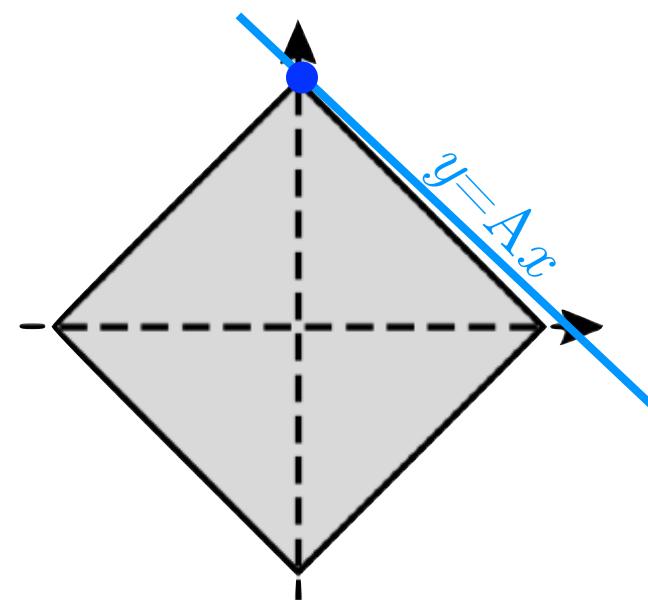
Relaxation

Solve $y = Ax$ where x is sparse

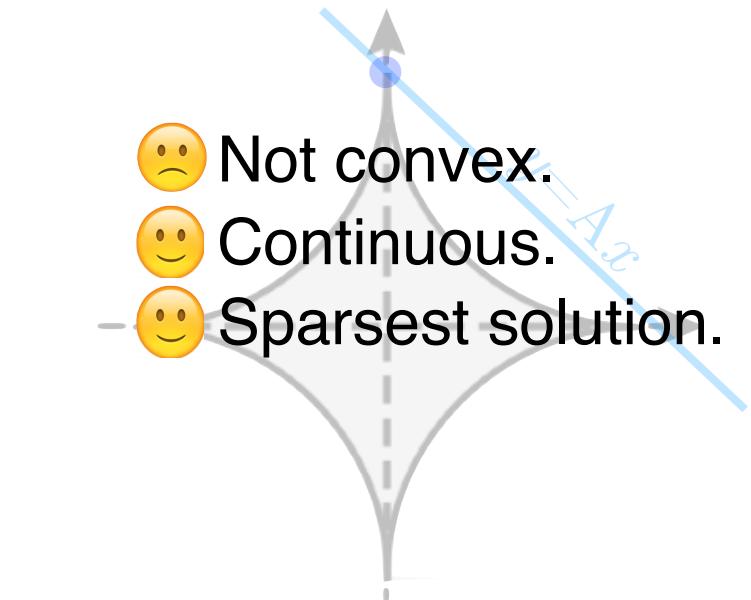
$$\min_{x \in \mathbb{R}^n} \|x\|_0 \text{ s.t. } y = Ax$$



$$\min_{x \in \mathbb{R}^n} \|x\|_1 \text{ s.t. } y = Ax$$

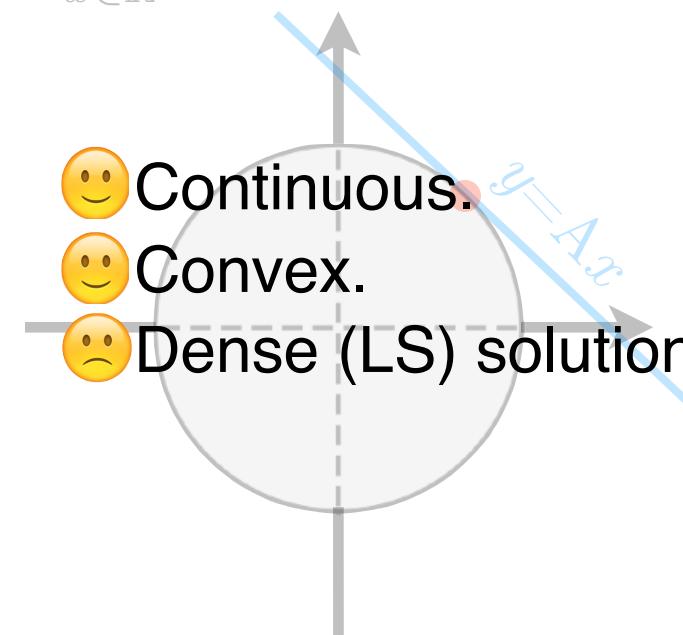


$$\min_{x \in \mathbb{R}^n} \|x\|_{0.5} \text{ s.t. } y = Ax$$

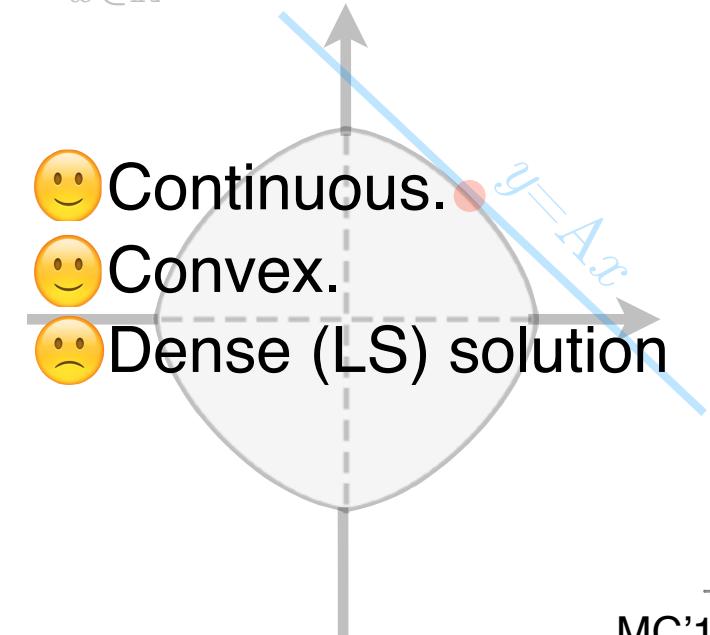


$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$$

$$\min_{x \in \mathbb{R}^n} \|x\|_2 \text{ s.t. } y = Ax$$



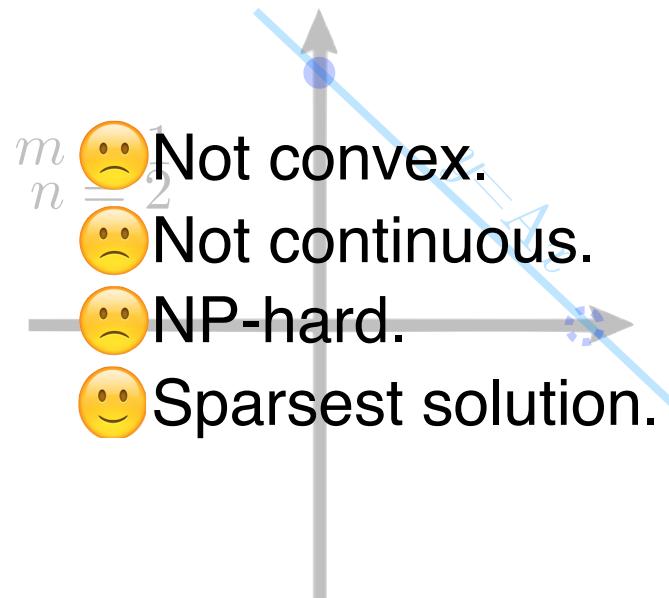
$$\min_{x \in \mathbb{R}^n} \|x\|_{1.5} \text{ s.t. } y = Ax$$



Relaxation

Solve $y = Ax$ where x is sparse

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \text{ s.t. } y = Ax$$



$$\min_{x \in \mathbb{R}^n} \|x\|_1 \text{ s.t. } y = Ax$$

- Basis Pursuit
- Continuous.
 - Convex.
 - Sparsest solution.

$$\min_{x \in \mathbb{R}^n} \|x\|_2 \text{ s.t. } y = Ax$$

-
- A graph showing the optimization problem $\min_{x \in \mathbb{R}^n} \|x\|_2$ subject to $y = Ax$. The horizontal axis is labeled y and the vertical axis is labeled x . A blue line represents the constraint $y = Ax$. The objective function is the ℓ_2 norm, represented by a circular region centered at the origin. A red dot marks the dense (LS) solution at the intersection of the axes.
- Continuous.
 - Convex.
 - Dense (LS) solution

$$\min_{x \in \mathbb{R}^n} \|x\|_{0.5} \text{ s.t. } y = Ax$$

-
- A graph showing the optimization problem $\min_{x \in \mathbb{R}^n} \|x\|_{0.5}$ subject to $y = Ax$. The horizontal axis is labeled y and the vertical axis is labeled x . A blue line represents the constraint $y = Ax$. The objective function is the $\ell_{0.5}$ norm, represented by a diamond shape centered at the origin. A blue dot marks the sparsest solution at the intersection of the axes.
- Not convex.
 - Continuous.
 - Sparsest solution.

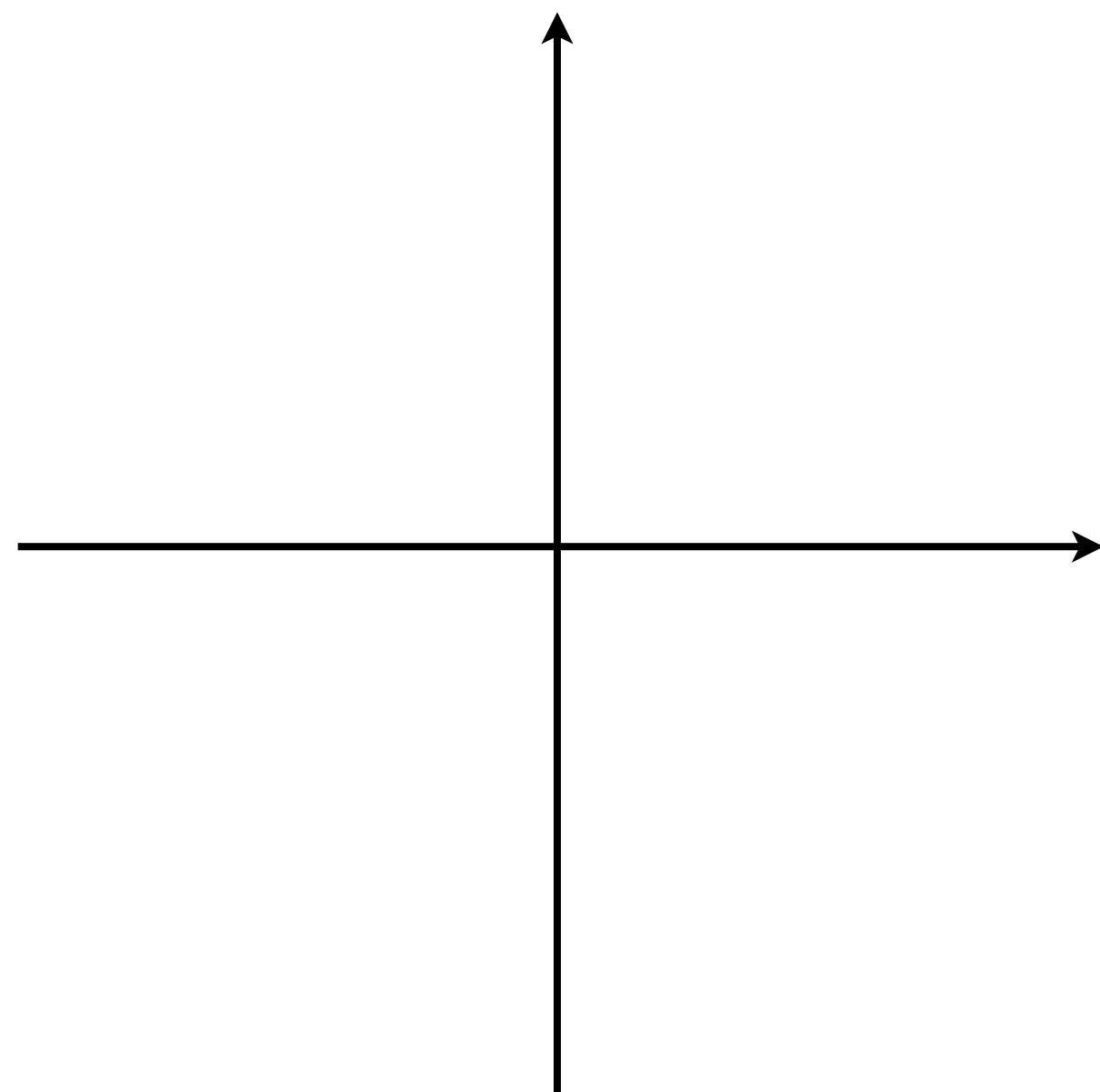
$$\min_{x \in \mathbb{R}^n} \|x\|_{1.5} \text{ s.t. } y = Ax$$

-
- A graph showing the optimization problem $\min_{x \in \mathbb{R}^n} \|x\|_{1.5}$ subject to $y = Ax$. The horizontal axis is labeled y and the vertical axis is labeled x . A blue line represents the constraint $y = Ax$. The objective function is the $\ell_{1.5}$ norm, represented by a circular region centered at the origin. A red dot marks the dense (LS) solution at the intersection of the axes.
- Continuous.
 - Convex.
 - Dense (LS) solution

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$$

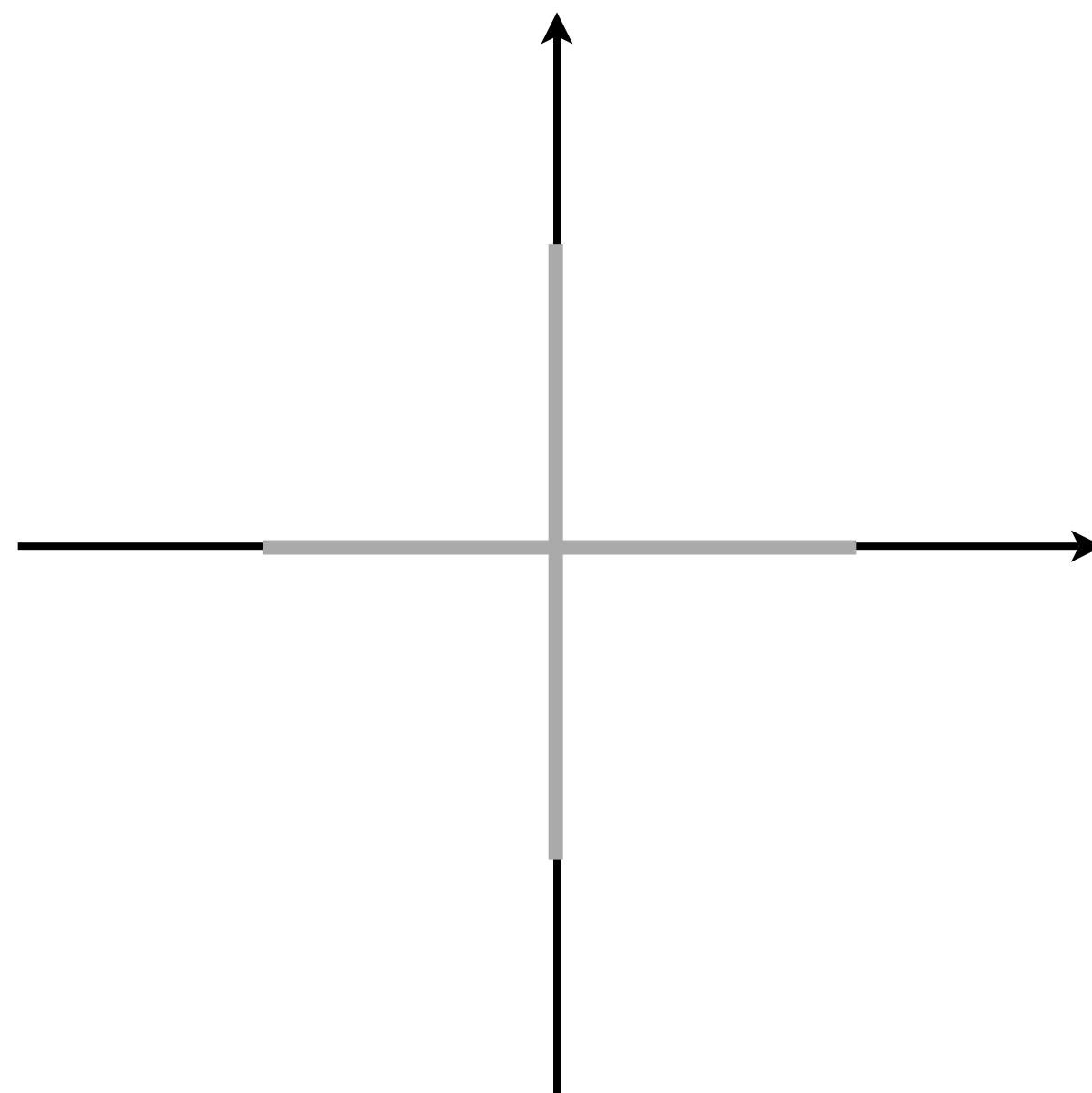
Tightest convex relaxation

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t. } y = Ax \quad (\text{BP})$$



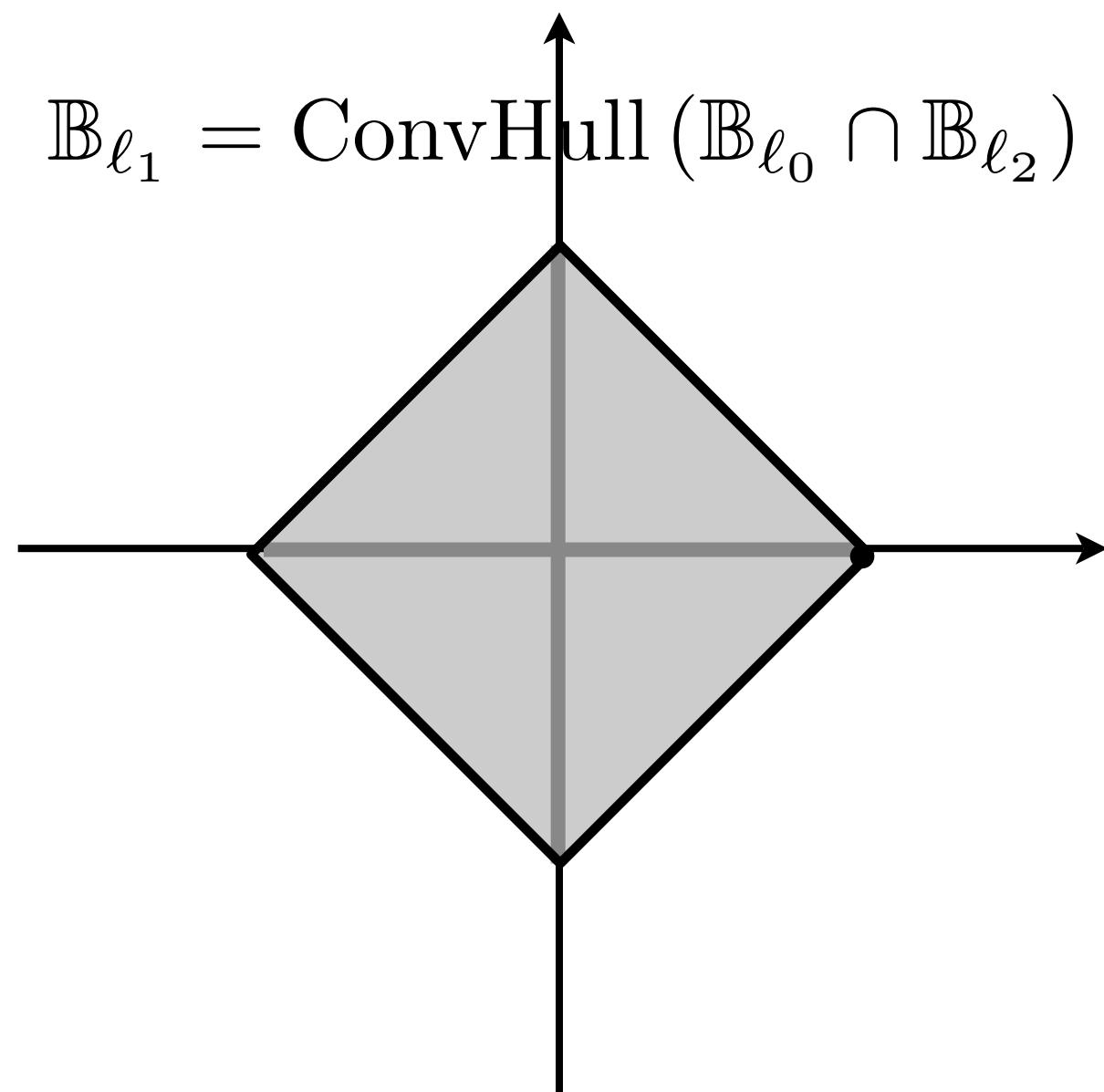
Tightest convex relaxation

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t. } y = Ax \quad (\text{BP})$$



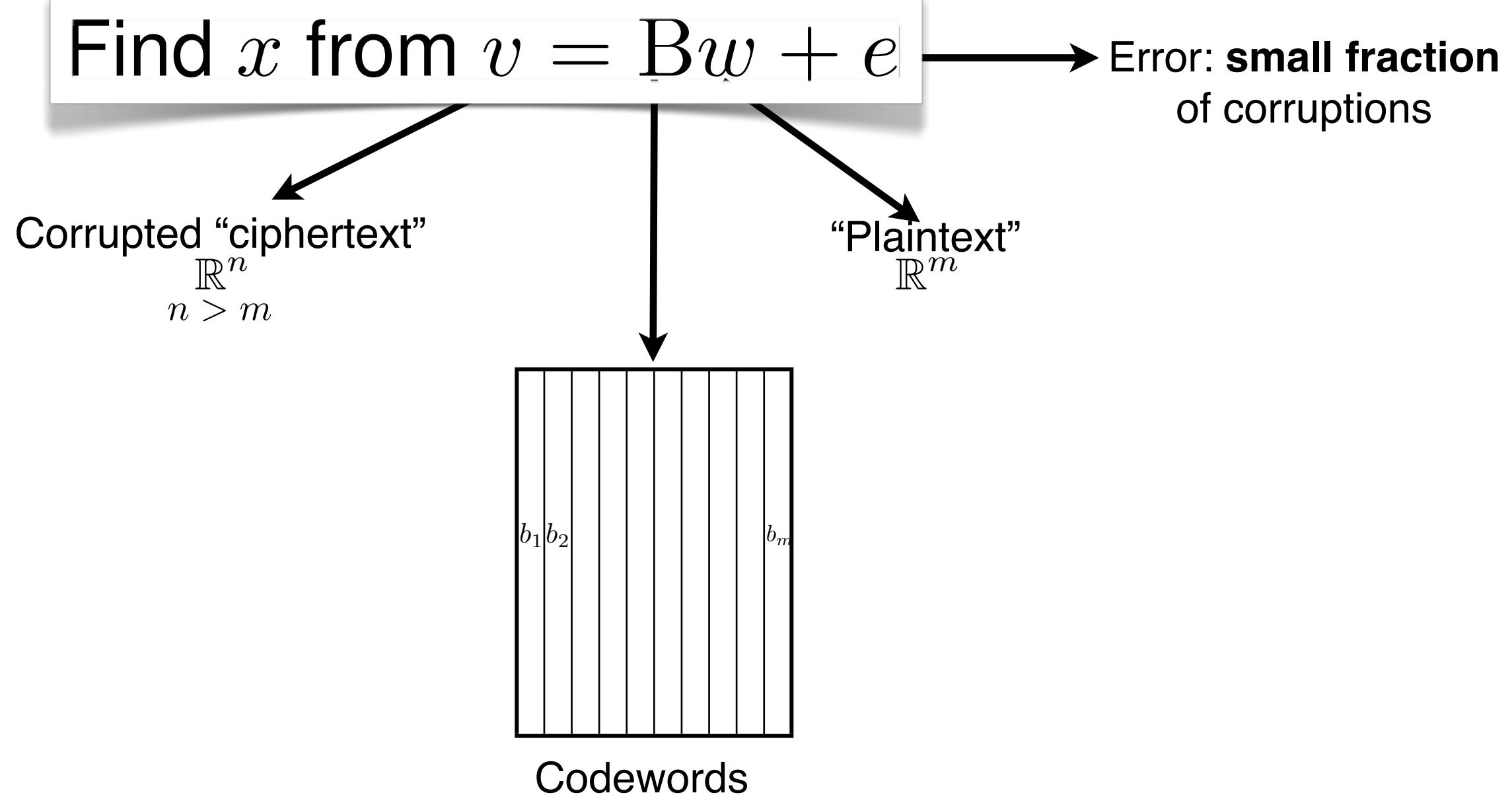
Tightest convex relaxation

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t. } y = Ax \quad (\text{BP})$$



ℓ_1 is the tightest convex relaxation of ℓ_0

Error correction problem



Error correction problem

Find x from $v = Bw + e$

- A such that $\text{span}(B) \subset \ker(A)$, i.e. $AB = 0$.
- Multiply v by A :

$$y = Av = Ae.$$

- Only a small fraction of corruptions means that e is sparse.
- The original problem can be cast as

$$x^* \in \underset{x \in \mathbb{R}^n}{\operatorname{Argmin}} \|x\|_1 \text{ s.t. } y = Ax$$

$$w^* = B^+ x^*$$

- Question : when $x^* = e$ so that $w^* = w$? (see following talks).

Optimization algorithms

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t. } y = Ax \quad (\text{BP})$$

- BP as a linear program :
- Decompose x in its positive and negative part and lift in \mathbb{R}^{2n} :

$$\min_{z \in \mathbb{R}^{2n}} \sum_{i=1}^2 z_i \quad \text{s.t. } y = [A \ - A]z, \ z \geq 0.$$

- Use your favourite LP solvers package : Cplex, Sedumi (IP), Mosek (IP), etc..
- Recover $x^* = (z_i^*)_{i=1}^n - (z_i^*)_{i=n+1}^{2n}$.
-  ● High accuracy.
-  ● Scaling with dimension n .
- BP as a linear program :
- Proximal splitting algorithms : DR, ADMM, Primal-Dual (MC of April 18th) :
-  ● Scale well with dimension : cost/iteration = $O(mn)$ vector/matrix multiplication and $O(n)$ soft-thresholding.
-  ● Iterative methods : less accurate.

Recovery guarantees

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t. } y = Ax \quad (\text{BP})$$

- Noiseless case $y = Ax_0$:
 - When (BP) has a unique solution that is the sparsest vector x_0 ?
 - Uniform guarantees : which conditions ensure recovery of *all* s -sparse signals ?
 - Non-uniform guarantees : which conditions ensure recovery of the s -sparse vector x_0 ?
 - Sample complexity bounds (random settings) : can we express the above conditions in terms of the problem dimensions (n, m, s) ? What are the optimal scalings ?
 - Necessary conditions ?
 - What if x_0 is only weakly sparse ?

Sensitivity/stability guarantees

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1, \quad \lambda > 0$$

(BP DN/LASSO)

- Noisy case $y = Ax_0 + \varepsilon$:
 - Study stability of (BP DN) solution(s) to the noise ε ?
 - ℓ_2 -stability :

Sensitivity/stability guarantees

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1, \quad \lambda > 0$$

(BPDN/LASSO)

- Noisy case $y = Ax_0 + \varepsilon$:
 - Study stability of (BP DN) solution(s) to the noise ε ?
 - ℓ_2 -stability:

Theorem (Typical statement) *Under conditions XX, and choice $\lambda = c \|\varepsilon\|_2$, there exists C such that any solution x^* of (BP DN) obeys*

$$\|x^* - x_0\| \leq C \|\varepsilon\|_2.$$

Sensitivity/stability guarantees

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1, \quad \lambda > 0$$

(BPDN/LASSO)

- Noisy case $y = Ax_0 + \varepsilon$:
 - Study stability of (BPDN) solution(s) to the noise ε ?
 - ℓ_2 -stability:

Theorem (Typical statement) *Under conditions XX, and choice $\lambda = c \|\varepsilon\|_2$, there exists C such that any solution x^* of (BPDN) obeys*

$$\|x^* - x_0\| \leq C \|\varepsilon\|_2.$$

- Support and sign stability (more stringent):

Theorem (Typical statement) *Under conditions XXXX, and choice $\lambda = f(\|\varepsilon\|_2, \min_i |x_i|)$, the unique solution x^* of (BPDN) obeys*

$$\text{supp}(x^*) = \text{supp}(x_0) \quad \text{and} \quad \text{sign}(x^*) = \text{sign}(x_0).$$

Sensitivity/stability guarantees

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1, \quad \lambda > 0$$

(BPDN/LASSO)

- Noisy case $y = Ax_0 + \varepsilon$:
 - Study stability of (BP DN) solution(s) to the noise ε ?
 - ℓ_2 -stability :

Theorem (Typical statement) *Under conditions XX, and choice $\lambda = c \|\varepsilon\|_2$, there exists C such that any solution x^* of (BP DN) obeys*

$$\|x^* - x_0\| \leq C \|\varepsilon\|_2.$$

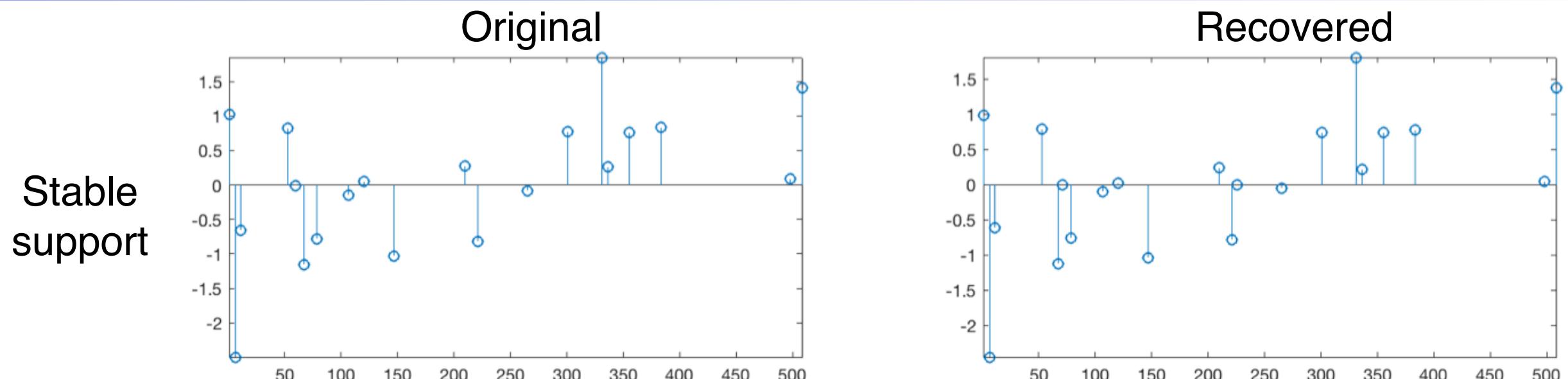
- Support and sign stability (more stringent) :

Theorem (Typical statement) *Under conditions XXXX, and choice $\lambda = f(\|\varepsilon\|_2, \min_i |x_i|)$, the unique solution x^* of (BP DN) obeys*

$$\text{supp}(x^*) = \text{supp}(x_0) \quad \text{and} \quad \text{sign}(x^*) = \text{sign}(x_0).$$

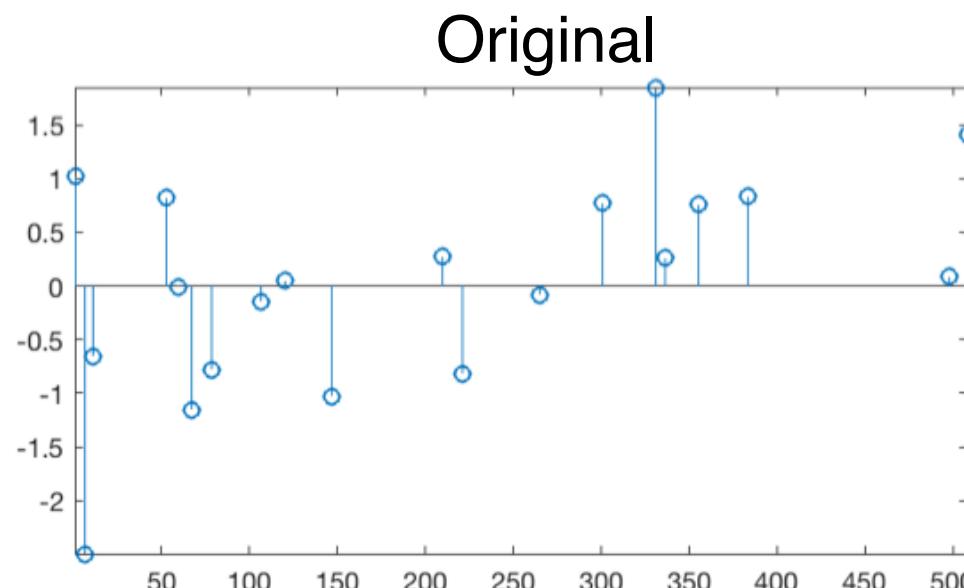
- Again uniform vs non-uniform guarantees.
- Sample complexity bounds (random setting) : can we express these conditions in terms of the problem dimensions (n, m, s) ? What are the scalings ?
- Necessary conditions ?
- What if x_0 is only weakly sparse ?

Sensitivity/stability guarantees

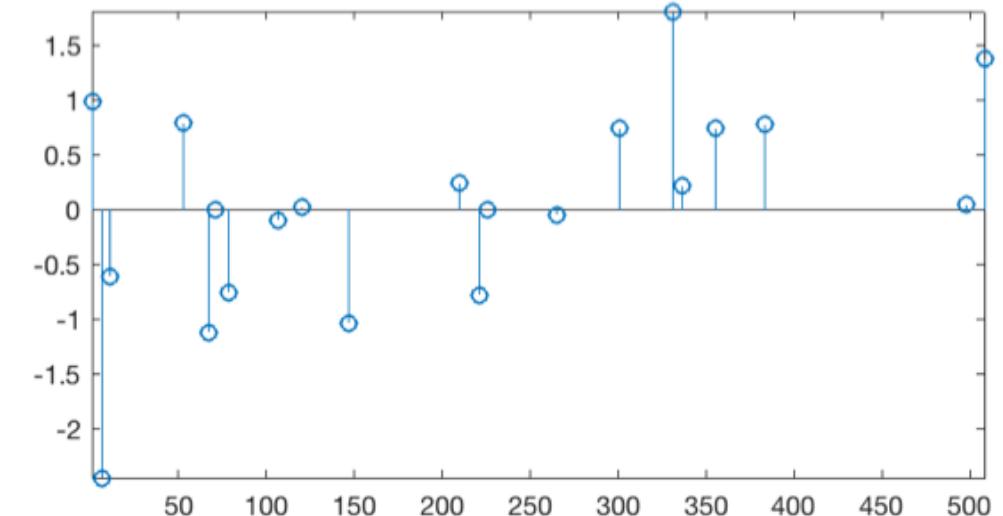


Sensitivity/stability guarantees

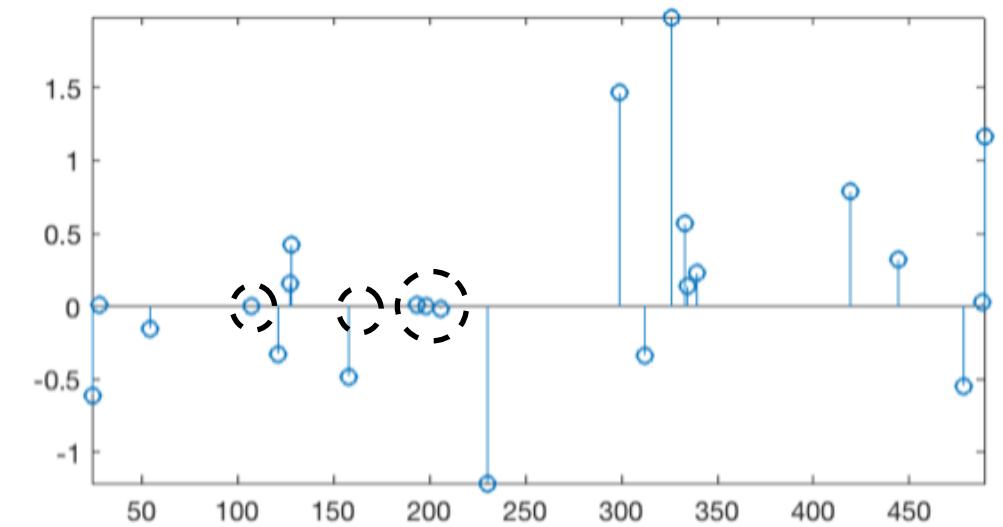
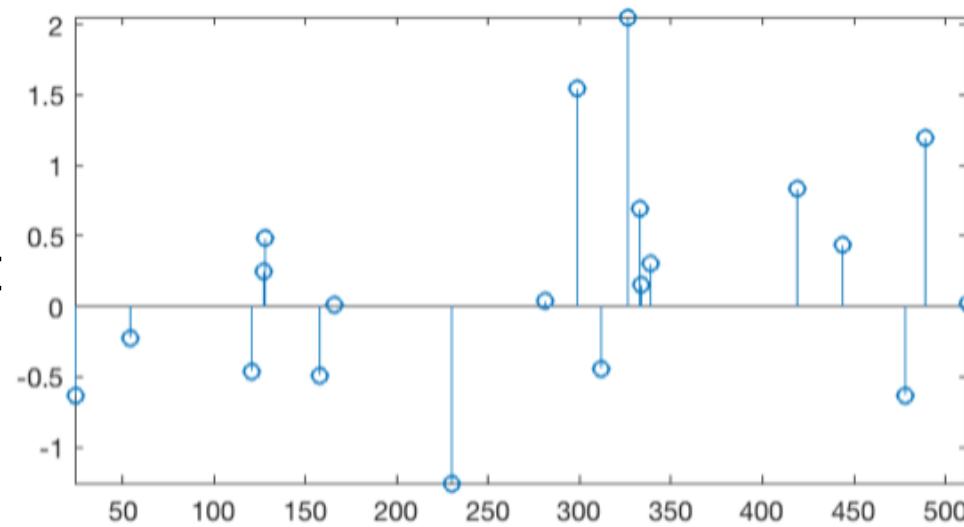
Stable support



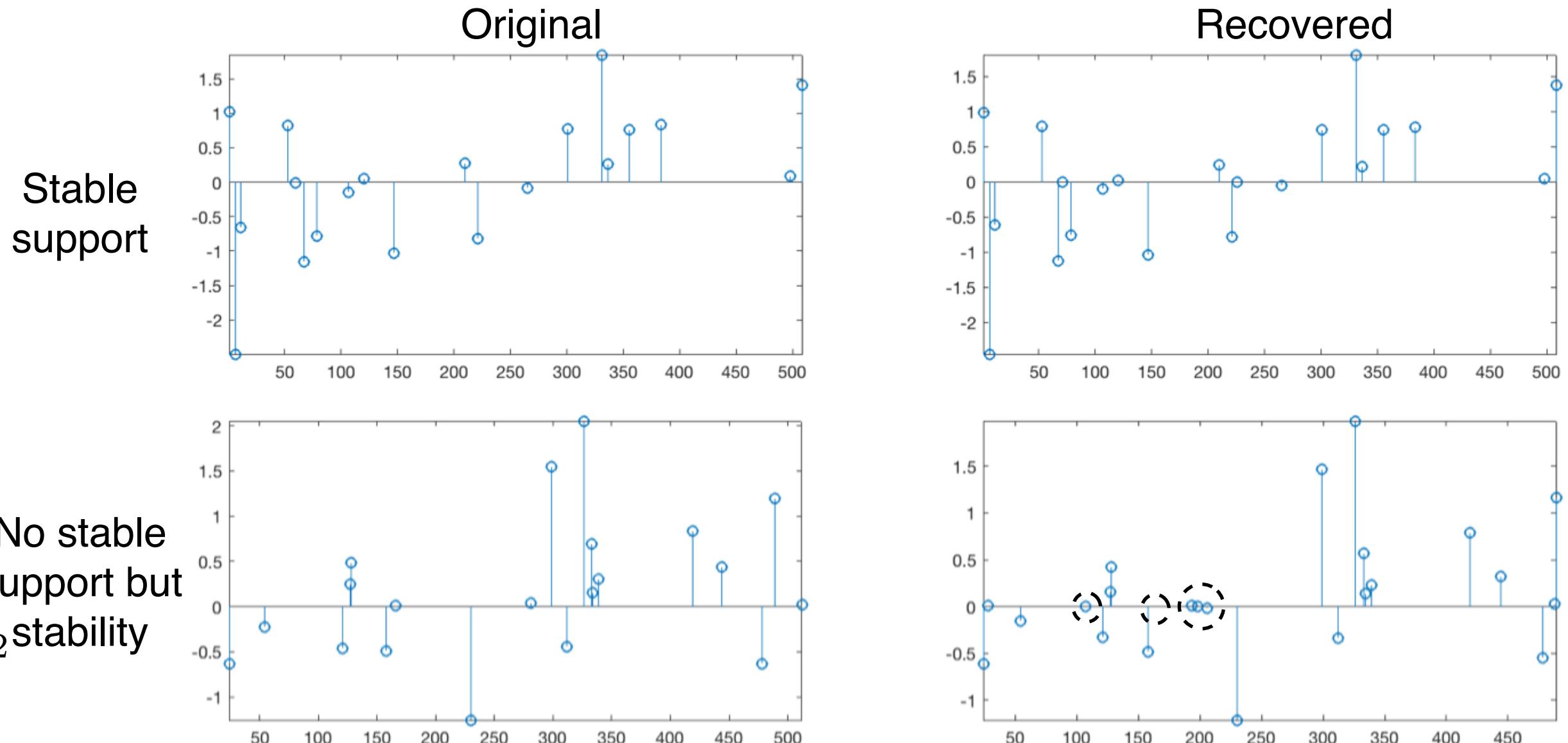
Recovered



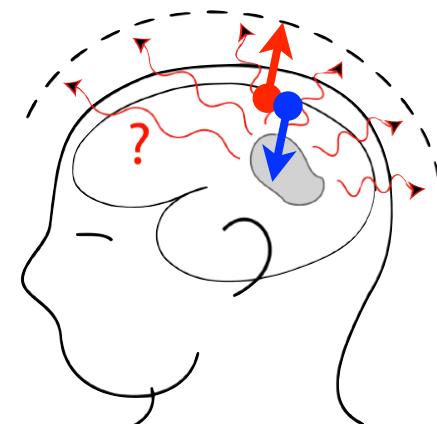
No stable support but ℓ_2 stability



Sensitivity/stability guarantees



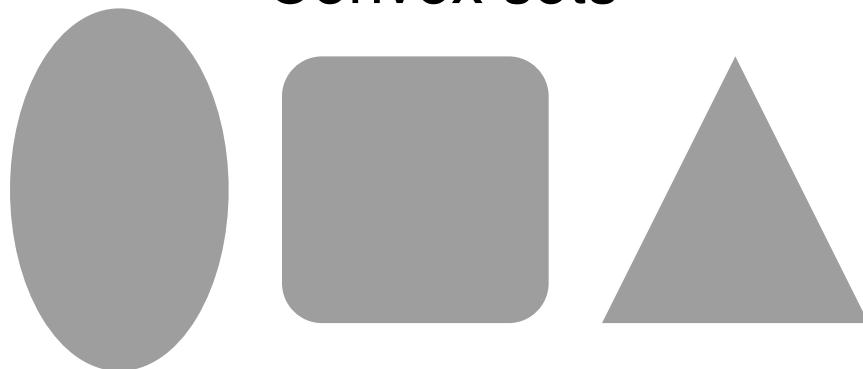
In some applications, what matters is stability of the support



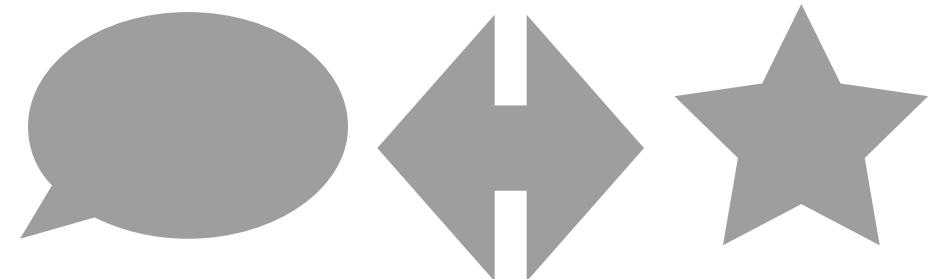
Guarantees from a geometrical perspective

Notions of convex analysis

Convex sets



Non-convex sets

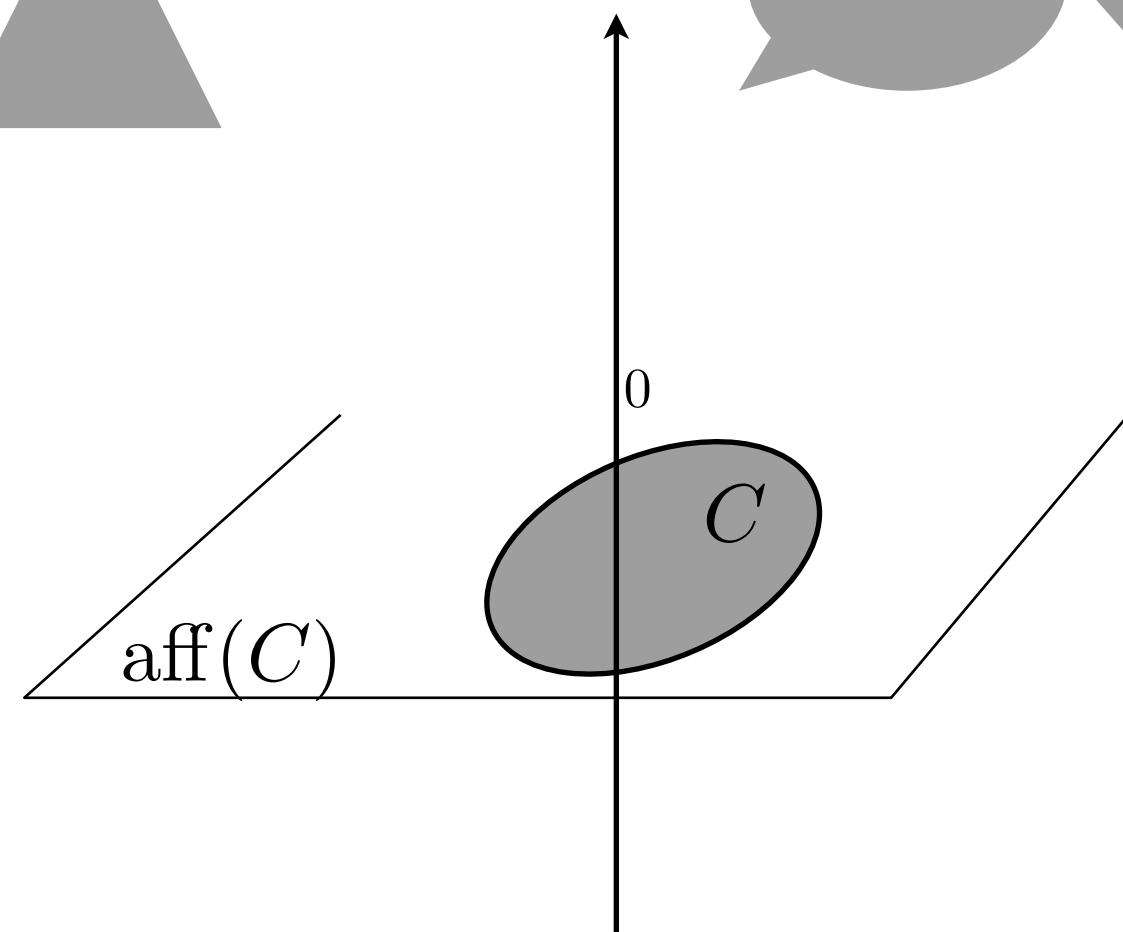
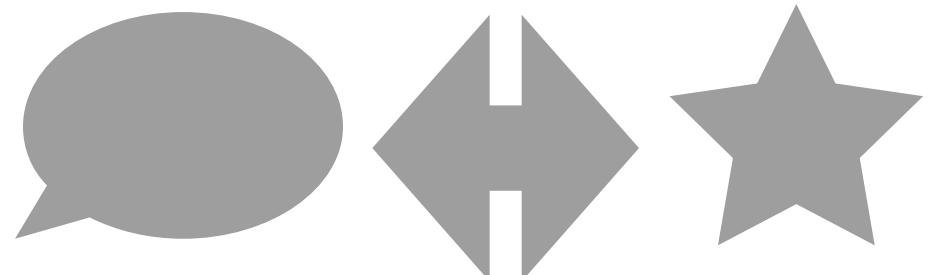


Notions of convex analysis

Convex sets



Non-convex sets

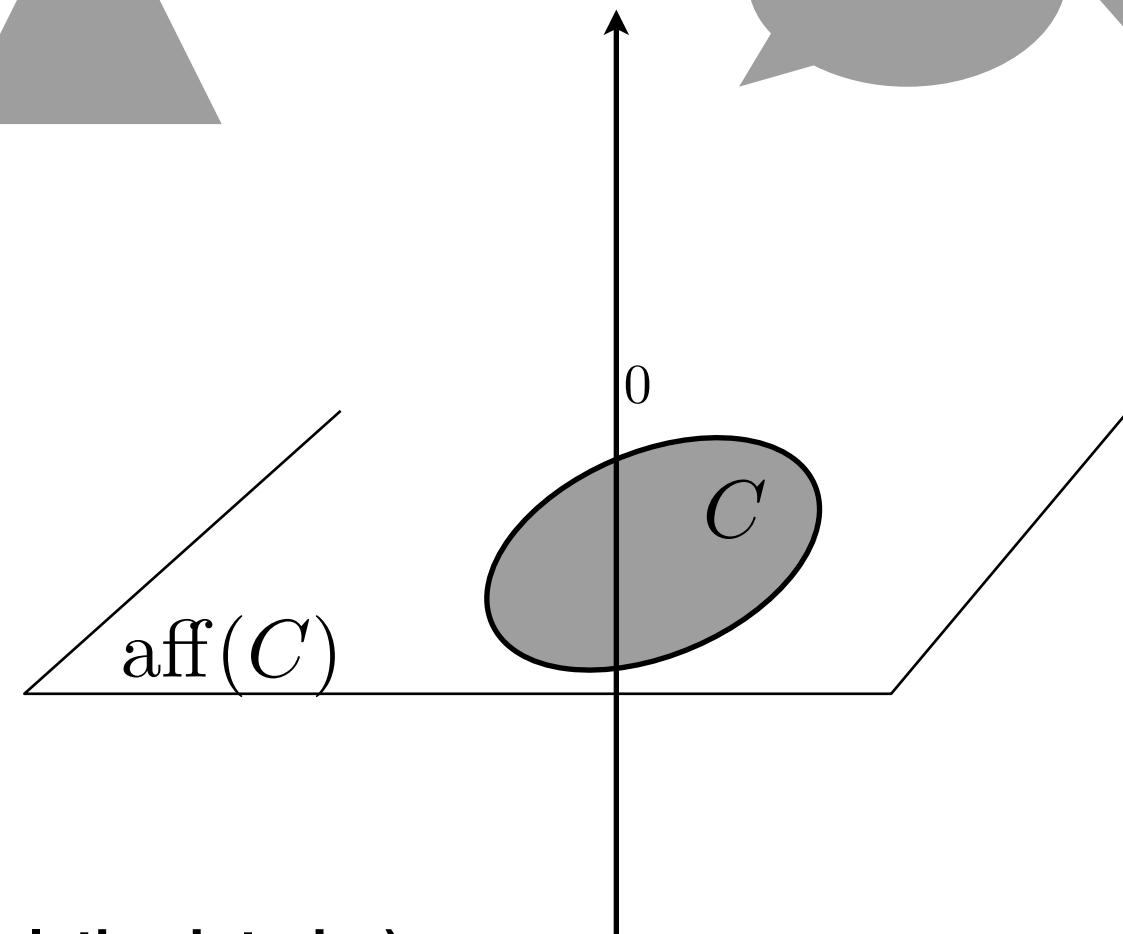
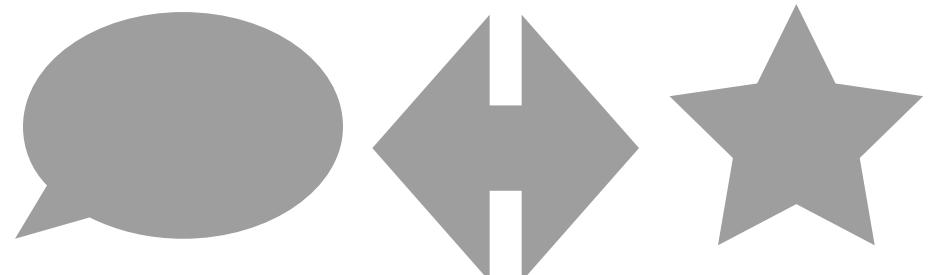


Notions of convex analysis

Convex sets



Non-convex sets



Definition (Relative interior)

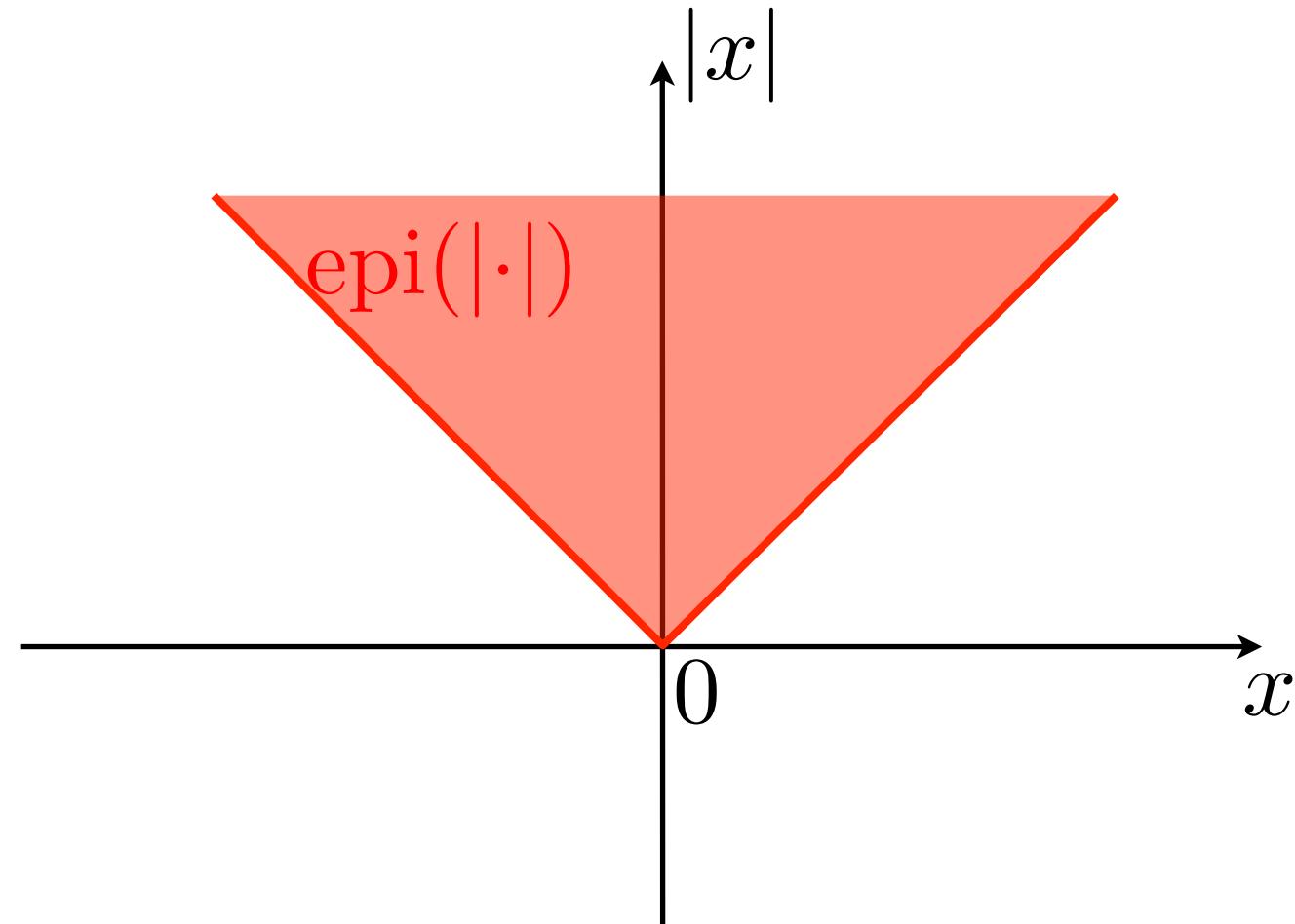
The relative interior $\text{ri}(C)$ of a convex set C is its interior relative to $\text{aff}(C)$.

C	$\text{aff}(C)$	$\text{ri}(C)$
$\{x\}$	$\{x\}$	$\{x\}$
$[x, x']$	line generated by (x, x')	$]x, x'[$
Simplex in \mathbb{R}^n	$\sum_{i=1}^n x_i = 1$	$\sum_{i=1}^n x_i = 1, x_i \in]0, 1[$

Subdifferential

Definition (Subdifferential) *The subdifferential of a convex function at $x \in \mathbb{R}^n$ is the set of slopes of affine functions minorizing f at x , i.e.*

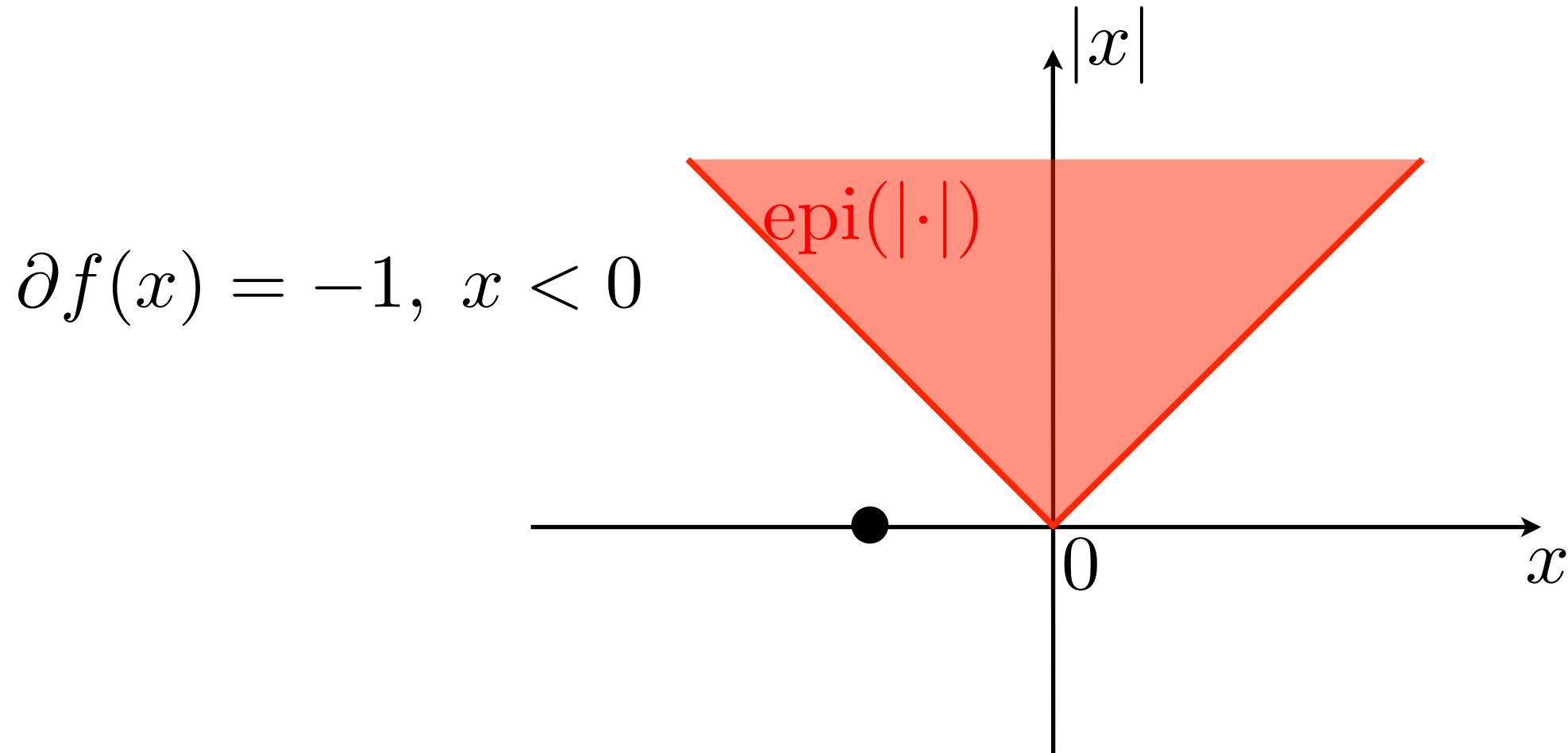
$$\partial f(x) = \{u \in \mathbb{R}^n : \forall z \in \mathbb{R}^n, f(z) \geq f(x) + \langle u, z - x \rangle\}.$$



Subdifferential

Definition (Subdifferential) *The subdifferential of a convex function at $x \in \mathbb{R}^n$ is the set of slopes of affine functions minorizing f at x , i.e.*

$$\partial f(x) = \{u \in \mathbb{R}^n : \forall z \in \mathbb{R}^n, f(z) \geq f(x) + \langle u, z - x \rangle\}.$$

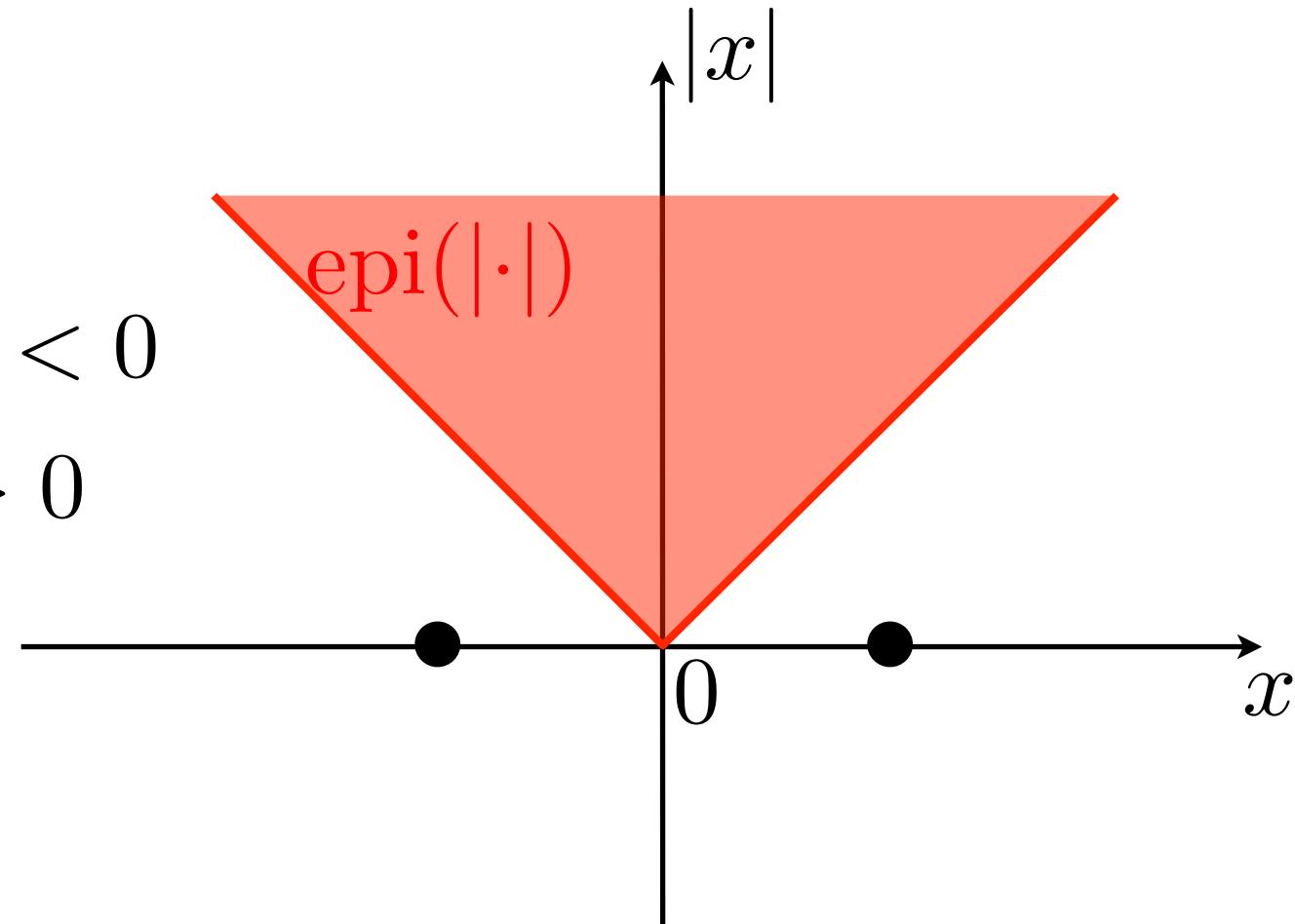


Subdifferential

Definition (Subdifferential) *The subdifferential of a convex function at $x \in \mathbb{R}^n$ is the set of slopes of affine functions minorizing f at x , i.e.*

$$\partial f(x) = \{u \in \mathbb{R}^n : \forall z \in \mathbb{R}^n, f(z) \geq f(x) + \langle u, z - x \rangle\}.$$

$$\begin{aligned}\partial f(x) &= -1, \quad x < 0 \\ \partial f(x) &= 1, \quad x > 0\end{aligned}$$



Subdifferential

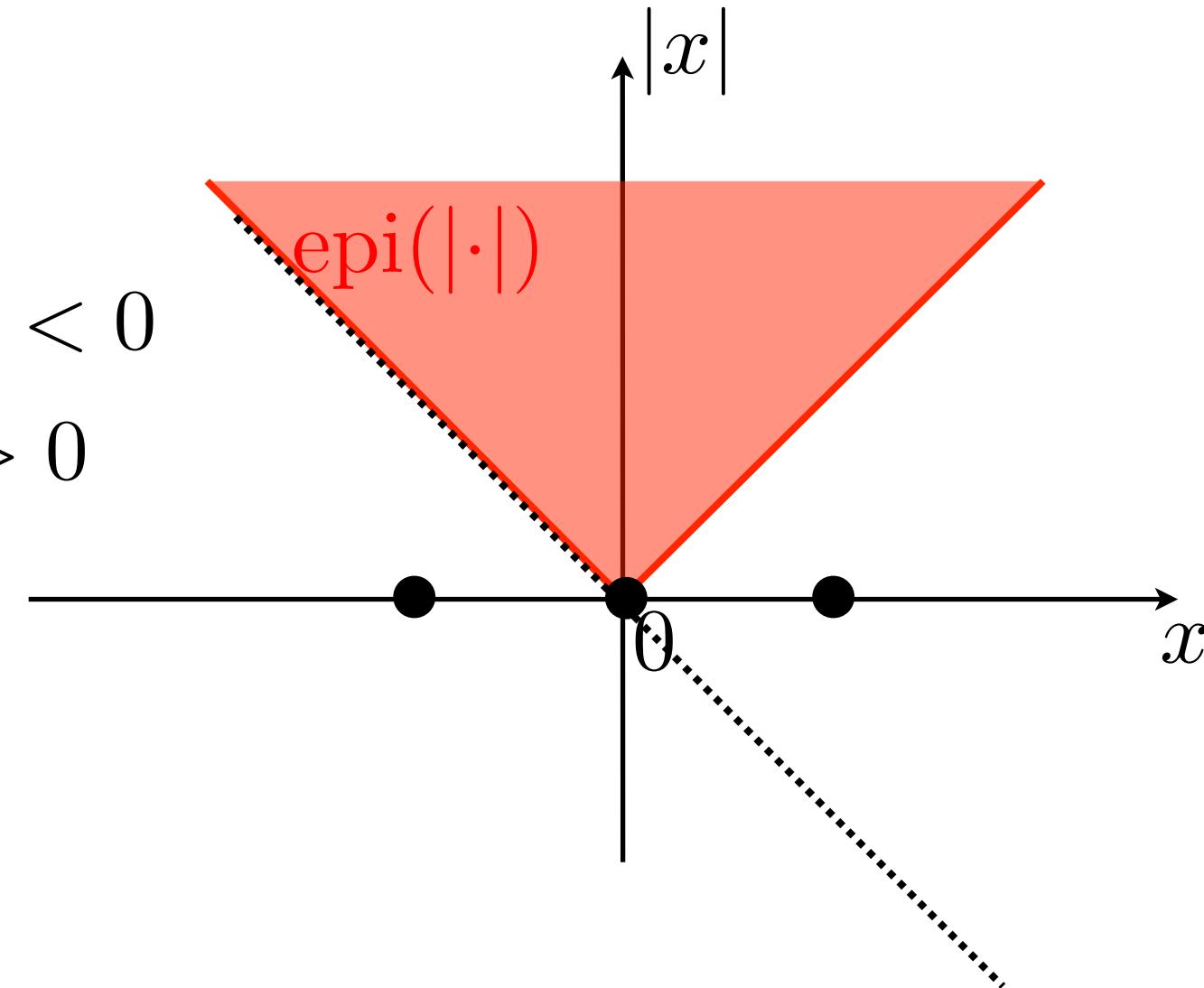
Definition (Subdifferential) *The subdifferential of a convex function at $x \in \mathbb{R}^n$ is the set of slopes of affine functions minorizing f at x , i.e.*

$$\partial f(x) = \{u \in \mathbb{R}^n : \forall z \in \mathbb{R}^n, f(z) \geq f(x) + \langle u, z - x \rangle\}.$$

$$\partial f(x) = -1, \quad x < 0$$

$$\partial f(x) = 1, \quad x > 0$$

$$\partial f(0) = [-1, 1]$$



Subdifferential

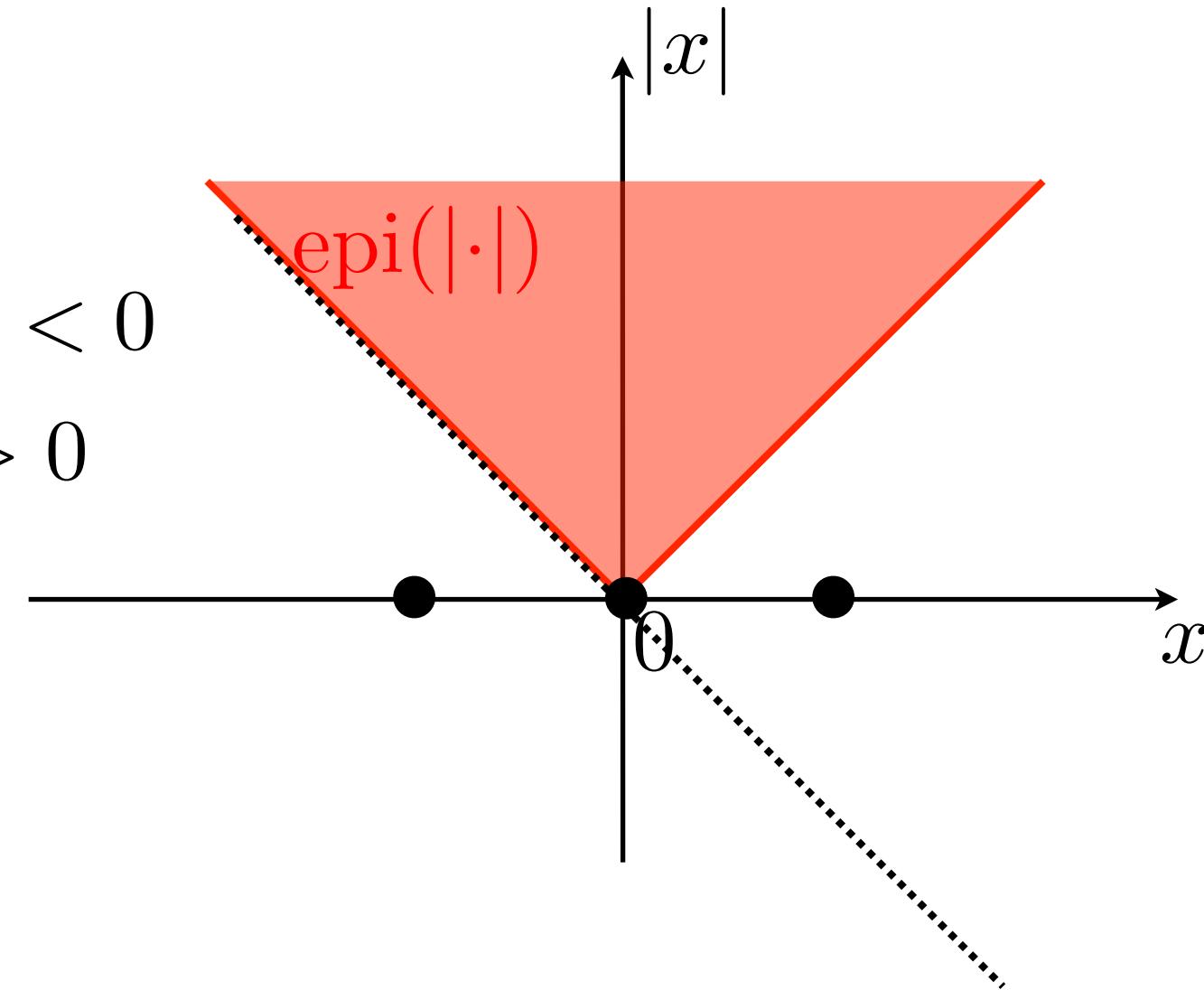
Definition (Subdifferential) The subdifferential of a convex function at $x \in \mathbb{R}^n$ is the set of slopes of affine functions minorizing f at x , i.e.

$$\partial f(x) = \{u \in \mathbb{R}^n : \forall z \in \mathbb{R}^n, f(z) \geq f(x) + \langle u, z - x \rangle\}.$$

$$\partial f(x) = -1, \quad x < 0$$

$$\partial f(x) = 1, \quad x > 0$$

$$\partial f(0) = [-1, 1]$$



$$\partial \|x\|_1 = \bigtimes_{i=1}^n \partial|x_i| \quad I \stackrel{\text{def}}{=} \text{supp}(x)$$

$$= \{u \in \mathbb{R}^n : u_I = \text{sign}(x_I), \|u\|_\infty \leq 1\}.$$

Normal cone

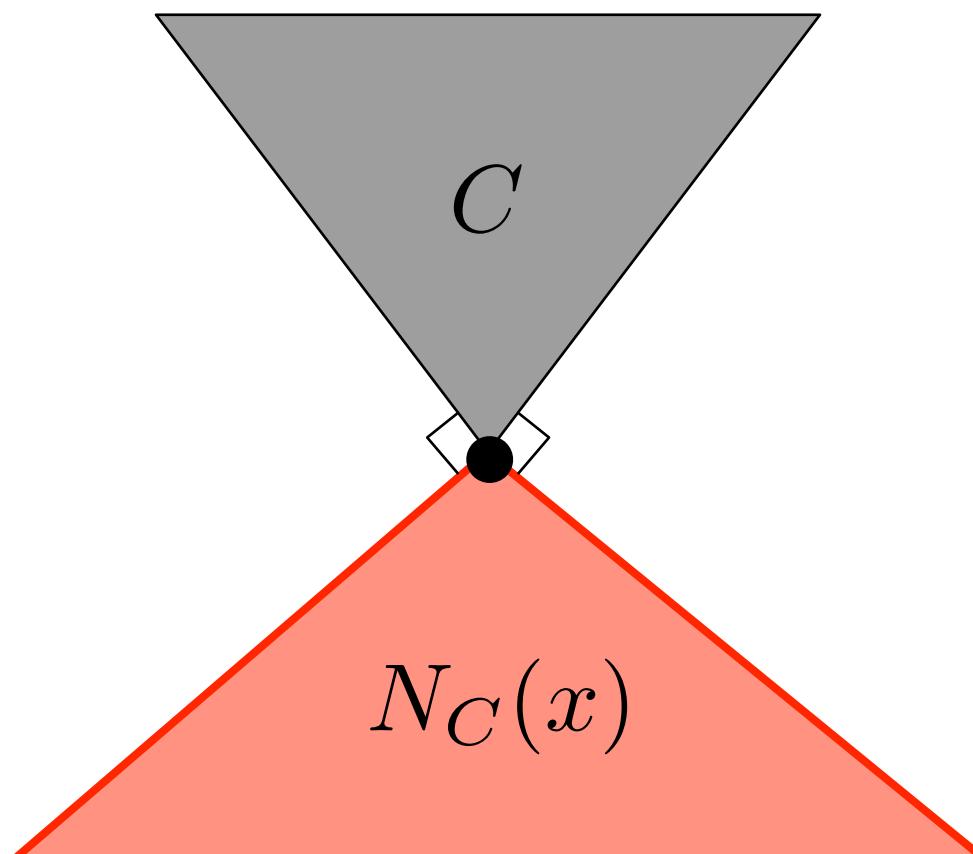
Definition (Normal cone) *The normal cone to a set C at $x \in C$ is*

$$N_C(x) = \{u \in \mathbb{R}^n : \langle u, z - x \rangle \leq 0, \forall z \in C\}.$$

Normal cone

Definition (Normal cone) *The normal cone to a set C at $x \in C$ is*

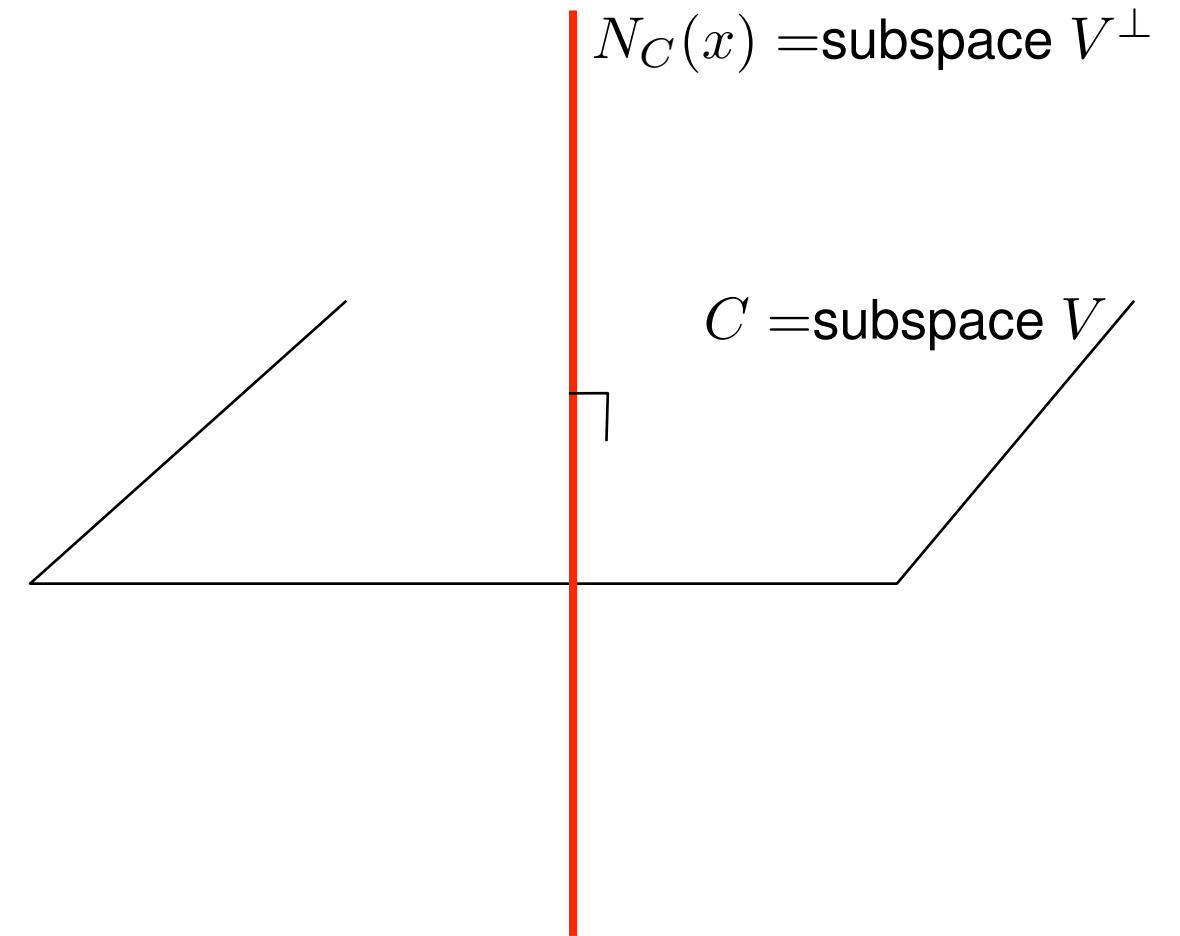
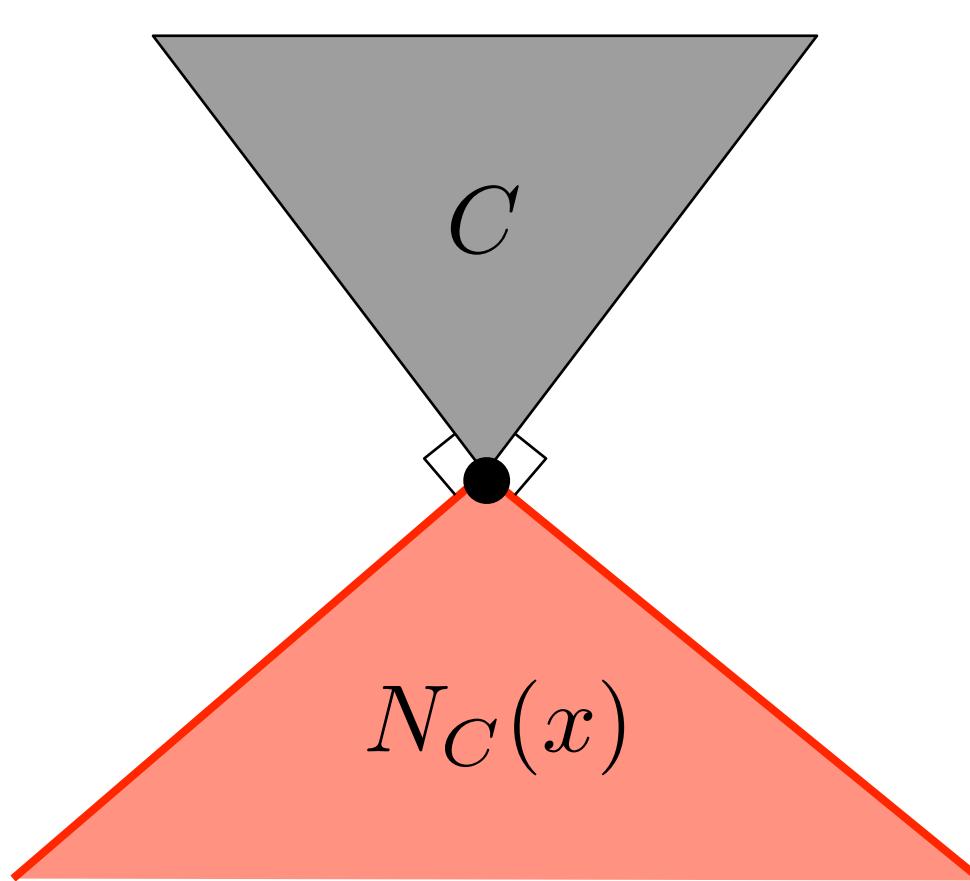
$$N_C(x) = \{u \in \mathbb{R}^n : \langle u, z - x \rangle \leq 0, \forall z \in C\}.$$



Normal cone

Definition (Normal cone) *The normal cone to a set C at $x \in C$ is*

$$N_C(x) = \{u \in \mathbb{R}^n : \langle u, z - x \rangle \leq 0, \forall z \in C\}.$$



Optimality conditions for (BP)

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t. } y = Ax \quad (\text{BP})$$

Optimality conditions for (BP)

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t. } y = Ax \quad (\text{BP})$$

$$x^* \in \operatorname{Argmin}_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t. } y = Ax$$

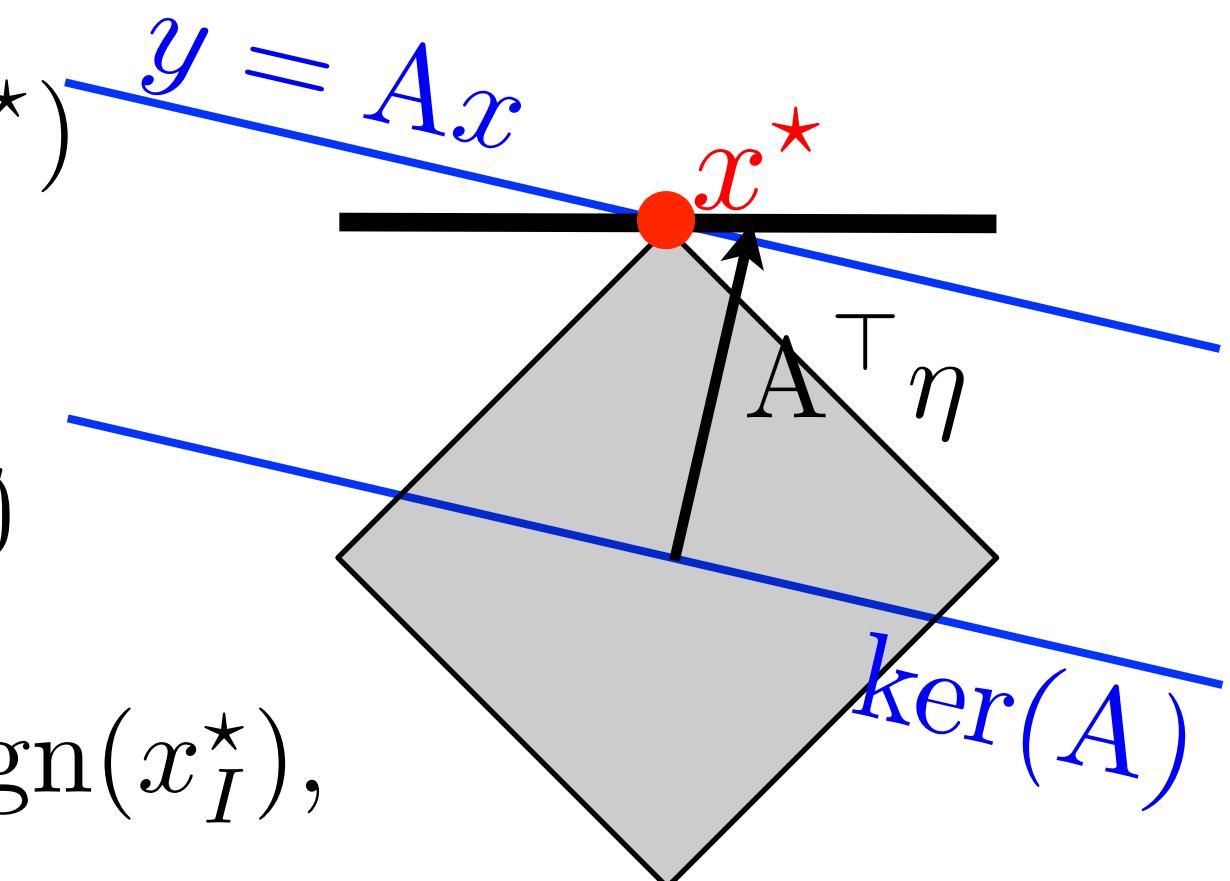
$$\iff 0 \in \partial \|x^*\|_1 + N_{\ker(A)}(x^*)$$

$$\iff 0 \in \partial \|x^*\|_1 + \operatorname{span}(A^\top)$$

$$\iff \operatorname{span}(A^\top) \cap \partial \|x^*\|_1 \neq \emptyset$$

$$\iff \exists \eta \in \mathbb{R}^m \text{ s.t. } \begin{cases} A_I^\top \eta = \operatorname{sign}(x_I^*), \\ \|A^\top \eta\|_\infty \leq 1. \end{cases}$$

[

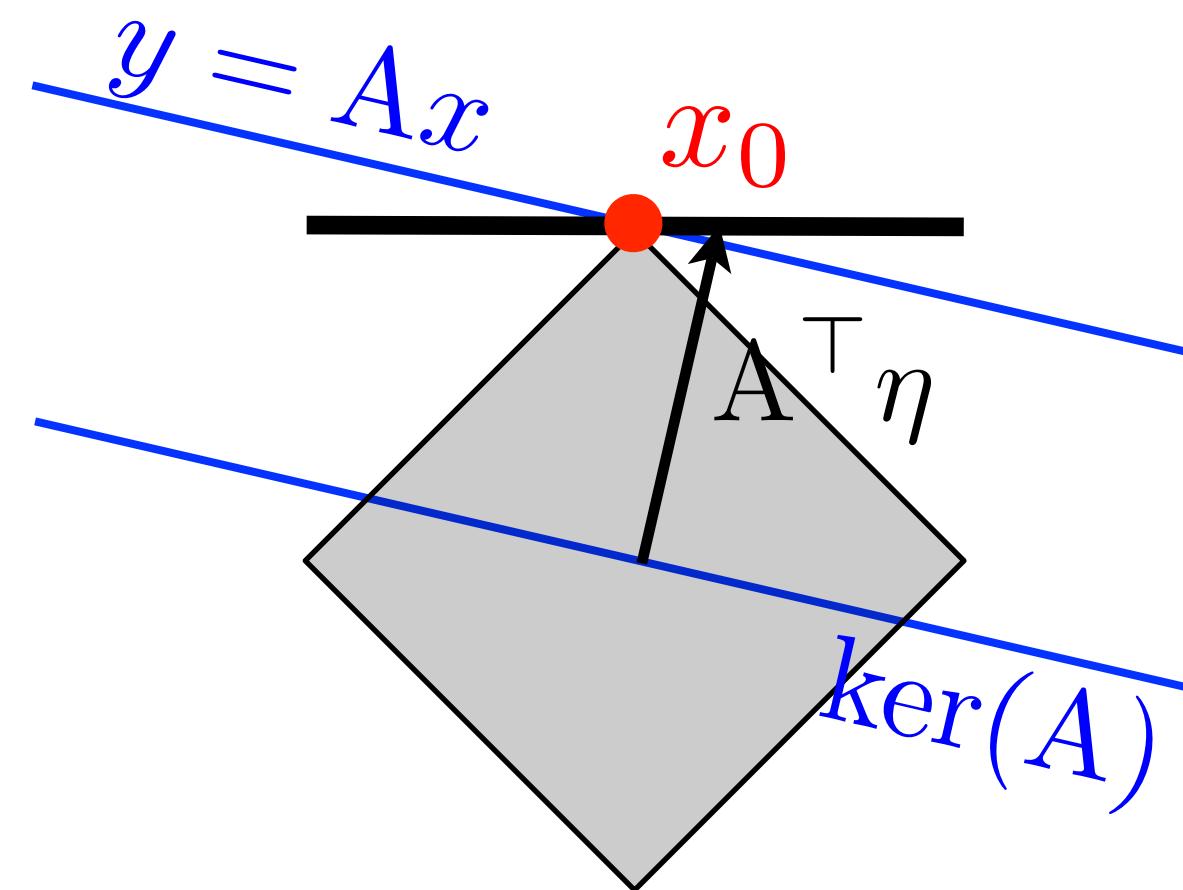


Dual certificate

Definition *The vector $\eta \in \mathbb{R}^m$ verifying the source condition*

$$A^\top \eta \in \partial \|x_0\|_1$$

*is called a **dual certificate** associated to x_0 .*



Non-degenerate dual certificate

Definition *The vector $\eta \in \mathbb{R}^m$ verifying the source condition*

$$\mathbf{A}^\top \eta \in \text{ri}(\partial \|x_0\|_1) \iff \mathbf{A}_I^\top \eta = \text{sign}((x_0)_I) \text{ and } \|\mathbf{A}_{I^c}^\top \eta\|_\infty < 1.$$

is called a non-degenerate dual certificate.

$$I \stackrel{\text{def}}{=} \text{supp}(x_0)$$

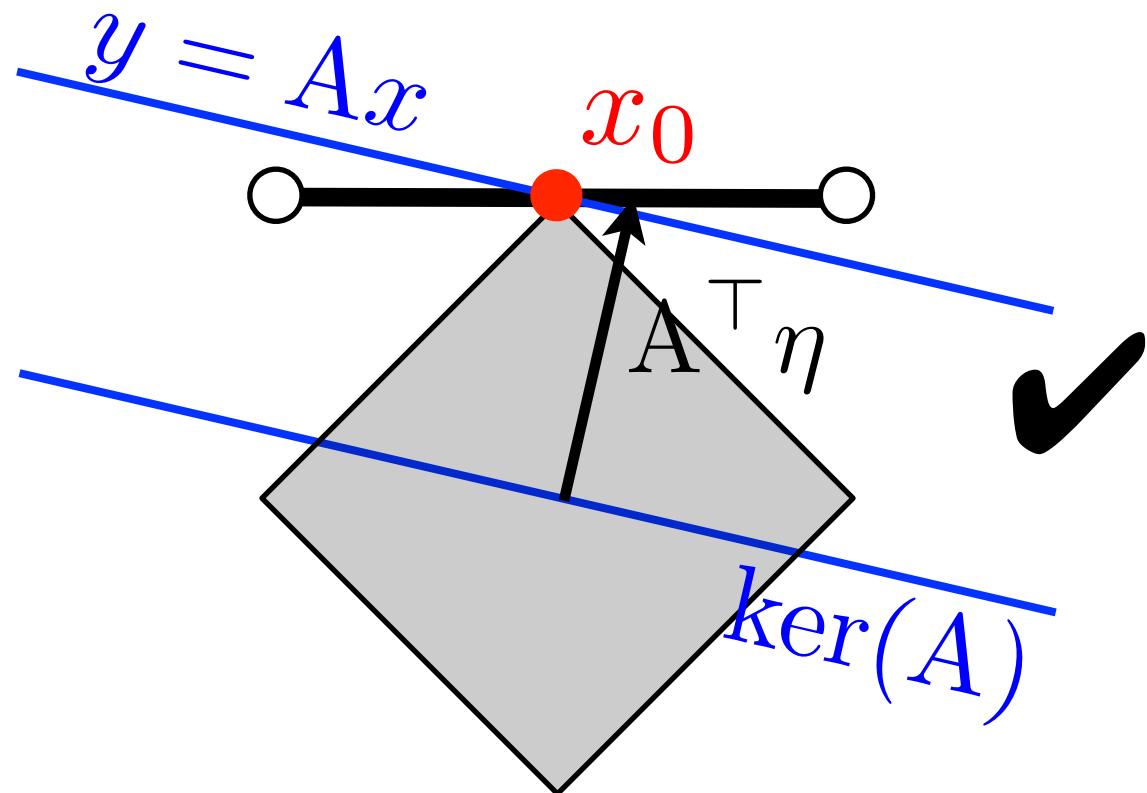
Non-degenerate dual certificate

Definition *The vector $\eta \in \mathbb{R}^m$ verifying the source condition*

$$A^\top \eta \in \text{ri}(\partial \|x_0\|_1) \iff A_I^\top \eta = \text{sign}((x_0)_I) \text{ and } \|A_{I^c}^\top \eta\|_\infty < 1.$$

is called a non-degenerate dual certificate.

$$I \stackrel{\text{def}}{=} \text{supp}(x_0)$$



Non-degenerate dual certificate

Definition The vector $\eta \in \mathbb{R}^m$ verifying the source condition

$$A^\top \eta \in \text{ri}(\partial \|x_0\|_1) \iff A_I^\top \eta = \text{sign}((x_0)_I) \text{ and } \|A_{I^c}^\top \eta\|_\infty < 1.$$

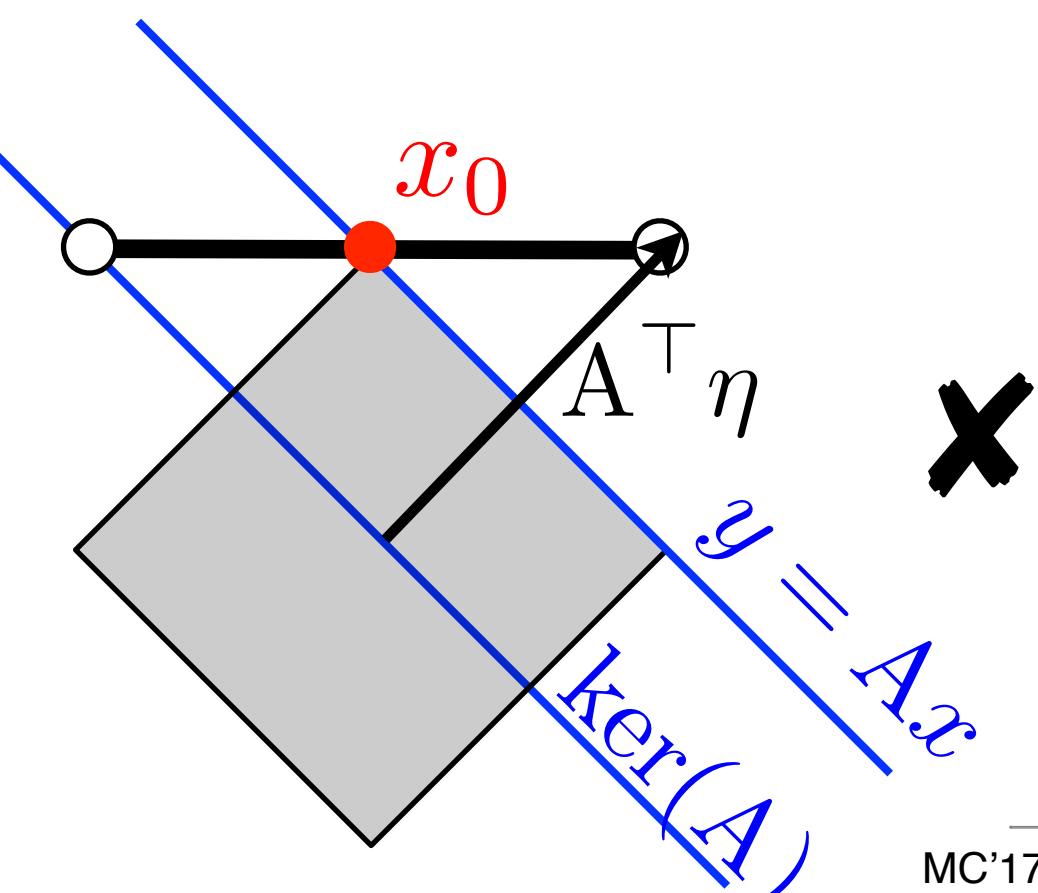
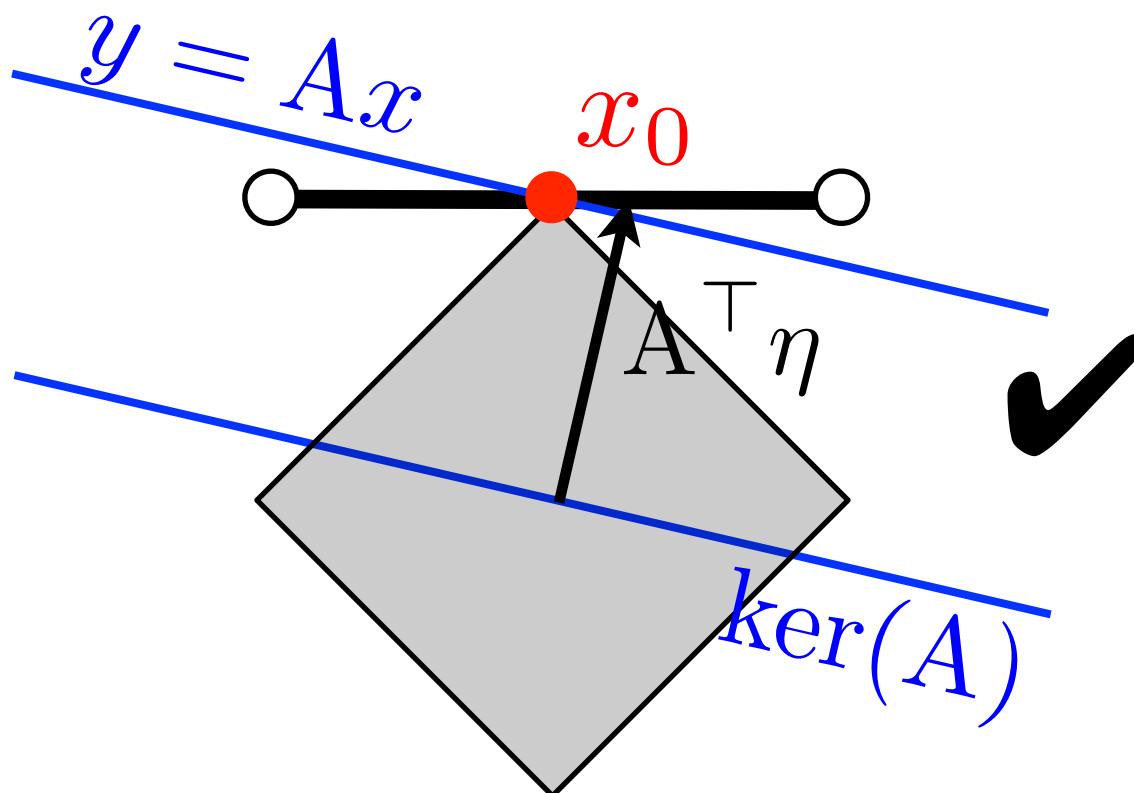
is called a **non-degenerate dual certificate**.

$$I \stackrel{\text{def}}{=} \text{supp}(x_0)$$

$A^\top \eta$ hits the relative boundary

$\iff \{x : y = Ax\}$ tangent to a higher dimensional face of x_0

\iff non-unique solution



Restricted Injectivity

Assumption A_I is full column rank, where $I \stackrel{\text{def}}{=} \text{supp}(x_0)$.

- A natural assumption.
- Assume noiseless case $y = Ax_0$
- Assume I is known, then

$$y = Ax_0 = A_I(x_0)_I.$$

- No hope to recover x_0 uniquely, even knowing its support, if A_I has a kernel.
- All recovery conditions in the literature assume a form of restricted injectivity.

Exact recovery

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t. } y = Ax \quad (\text{BP})$$

Theorem Let $I = \text{supp}(x_0)$. Assume that there exists a non-degenerate dual certificate at x_0 and A_I is full-rank. Then x_0 is the unique solution to (BP).

- Even necessary when x_0 is non-trivial.

Stability without support recovery

$$y = Ax_0 + \varepsilon$$

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1, \quad \lambda > 0$$

(BPDN/LASSO)

Theorem Let $I = \text{supp}(x_0)$. Assume that there exists a non-degenerate dual certificate η at x_0 and A_I is full-rank. Then, choosing $\lambda = c \|\varepsilon\|_2$, $c > 0$, any minimizer x^* of (BPDN/LASSO) obeys

$$\|x^* - x_0\|_2 \leq C(c, A, I, \eta) \|\varepsilon\|_2.$$

Stability without support recovery

$$y = Ax_0 + \varepsilon$$

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1, \quad \lambda > 0$$

(BPDN/LASSO)

Theorem Let $I = \text{supp}(x_0)$. Assume that there exists a non-degenerate dual certificate η at x_0 and A_I is full-rank. Then, choosing $\lambda = c \|\varepsilon\|_2$, $c > 0$, any minimizer x^* of (BPDN/LASSO) obeys

$$\|x^* - x_0\|_2 \leq C(c, A, I, \eta) \|\varepsilon\|_2.$$

- Even necessary when x_0 is non-trivial.

Stable support and sign recovery

$$y = Ax_0 + \varepsilon$$

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1, \quad \lambda > 0$$

(BPDN/LASSO)

Theorem Let $I = \text{supp}(x_0)$. Assume that A_I is full-rank and

$$\eta_F = A_I (A_I^\top A_I)^{-1} \text{sign}((x_0)_I)$$

is a non-degenerate dual certificate. Then, choosing

$$c_1 \|\varepsilon\|_2 < \lambda < c_2 \min_{i \in I} |(x_0)_i|,$$

(BPDN/LASSO) has a unique solution x^* which moreover satisfies

$$\text{supp}(x^*) = I \text{ and } \text{sign}(x^*) = \text{sign}(x_0).$$

Stable support and sign recovery

$$y = Ax_0 + \varepsilon$$

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1, \quad \lambda > 0$$

(BPDN/LASSO)

Theorem Let $I = \text{supp}(x_0)$. Assume that A_I is full-rank and

$$\eta_F = A_I (A_I^\top A_I)^{-1} \text{sign}((x_0)_I)$$

is a non-degenerate dual certificate. Then, choosing

$$c_1 \|\varepsilon\|_2 < \lambda < c_2 \min_{i \in I} |(x_0)_i|,$$

(BPDN/LASSO) has a unique solution x^* which moreover satisfies

$$\text{supp}(x^*) = I \text{ and } \text{sign}(x^*) = \text{sign}(x_0).$$

- Almost necessary when x_0 is non-trivial.

Take-away messages

- Convex relaxation is good for sparse recovery.
- Many (tight) guarantees with nice geometrical insight:
 - Exact noiseless recovery.
 - Stability without support recovery.
 - Stable support recovery.
- Can we translate these conditions into sample complexity bounds ?
- Yes: random measurements (Guillaume Lecué).

[**https://fadili.users.greyc.fr/**](https://fadili.users.greyc.fr/)

**Thanks
Any questions ?**