Sparsity and Compressed Sensing

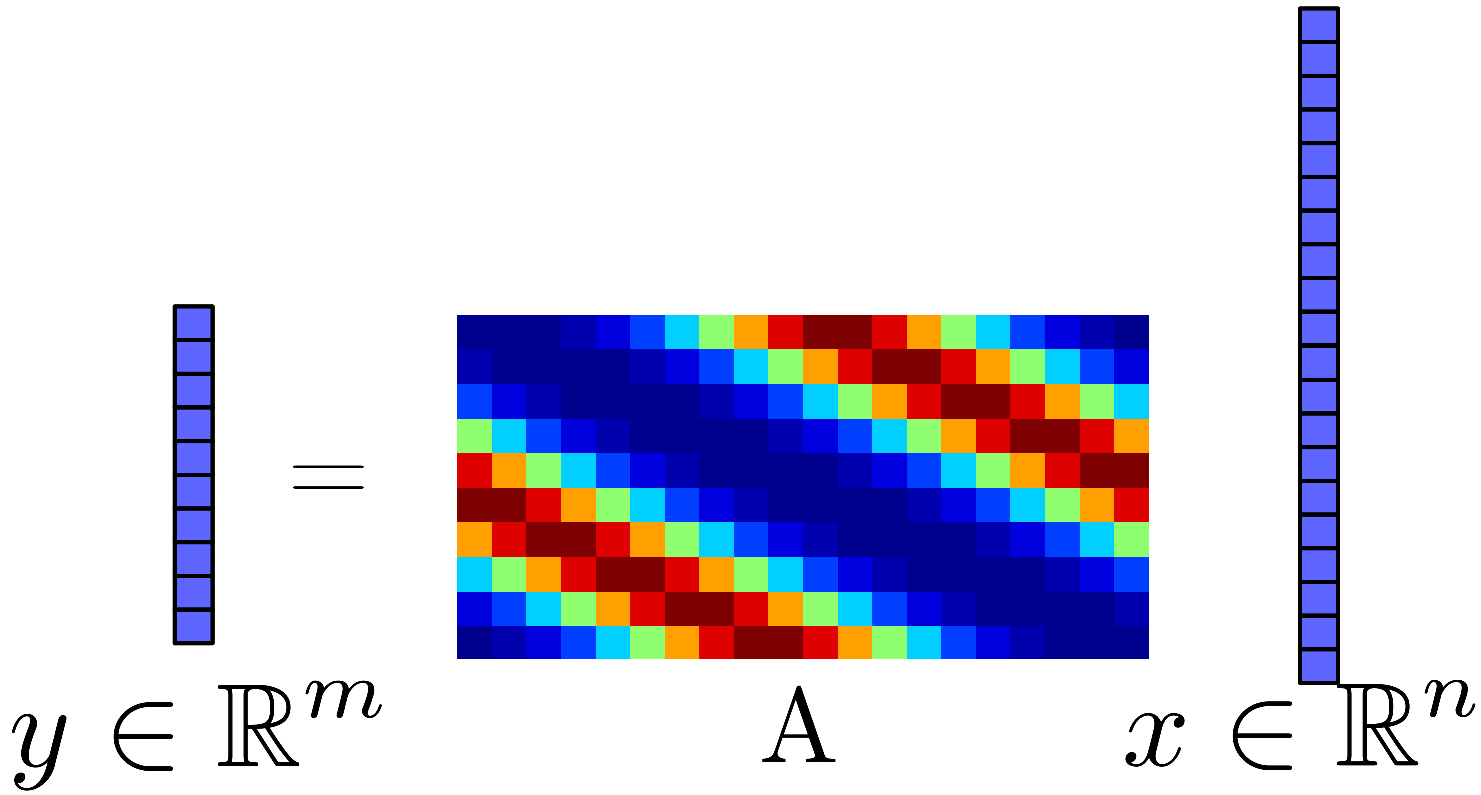
Jalal Fadili
Normandie Université-ENSICAEN, GREYC

Mathematical coffees 2017

Normandie Université

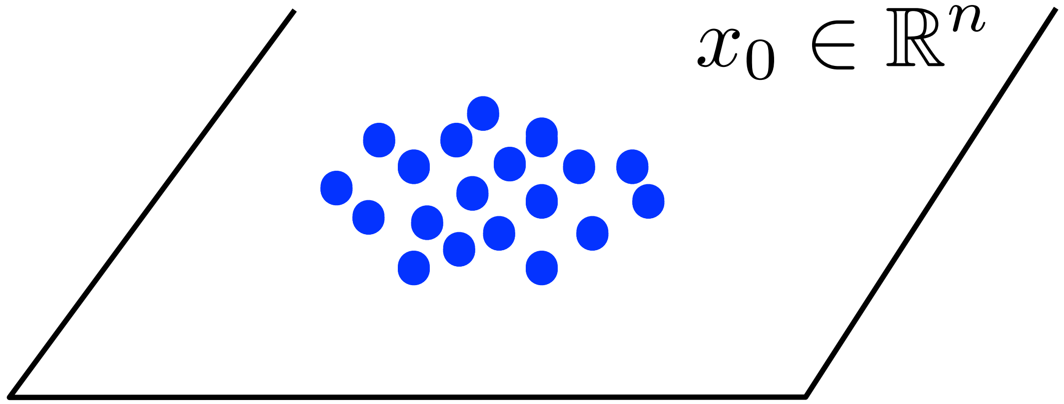
ENSI CAEN
ÉCOLE PUBLIQUE D'INGÉNIEURS
CENTRE DE RECHERCHE

Recap: linear inverse problems

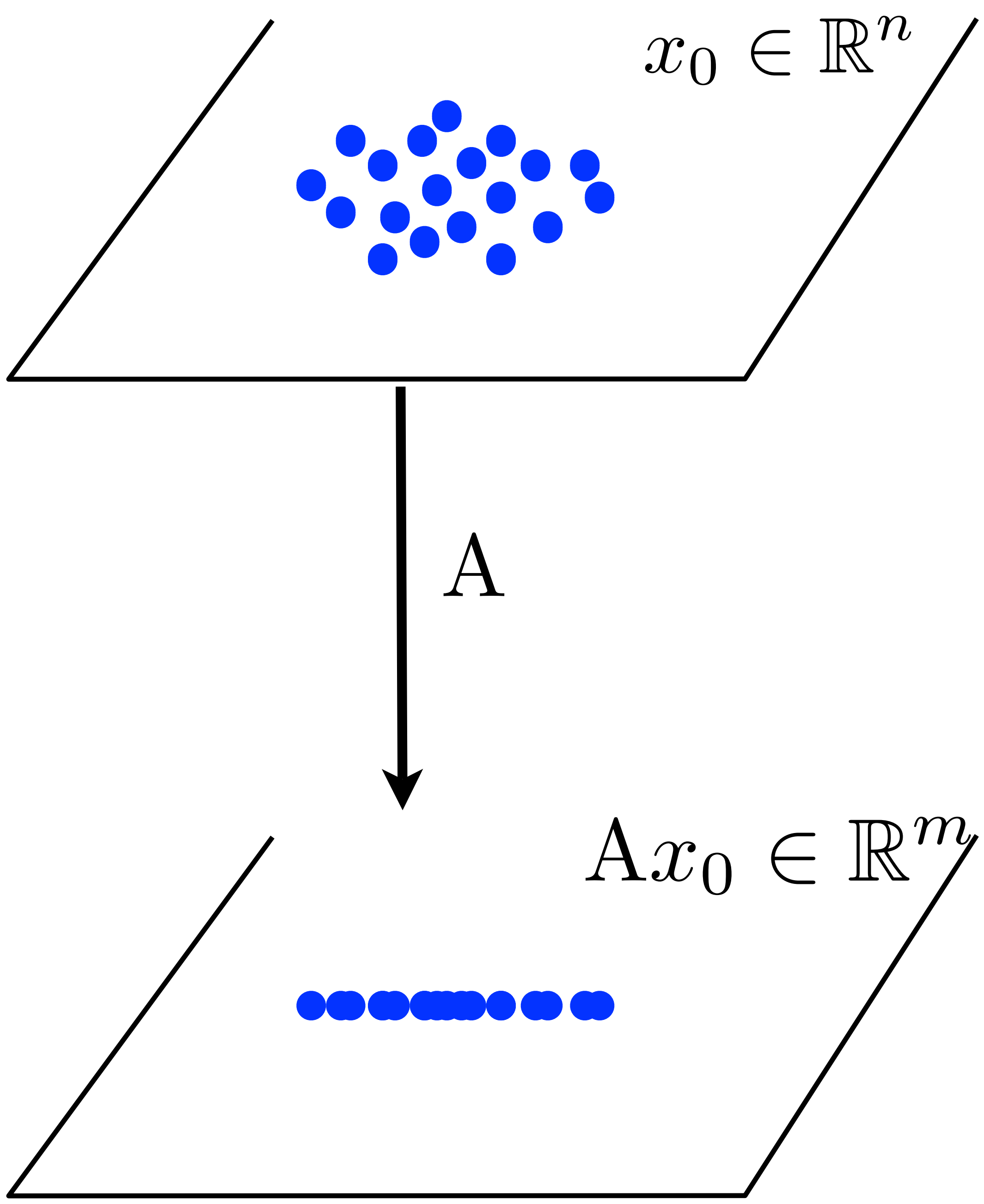


- A is fat (underdetermined system).
- Solution is not unique (fundamental theorem of linear algebra).
- Are we stuck ?
- No if the dimension of x is **intrinsically small**.

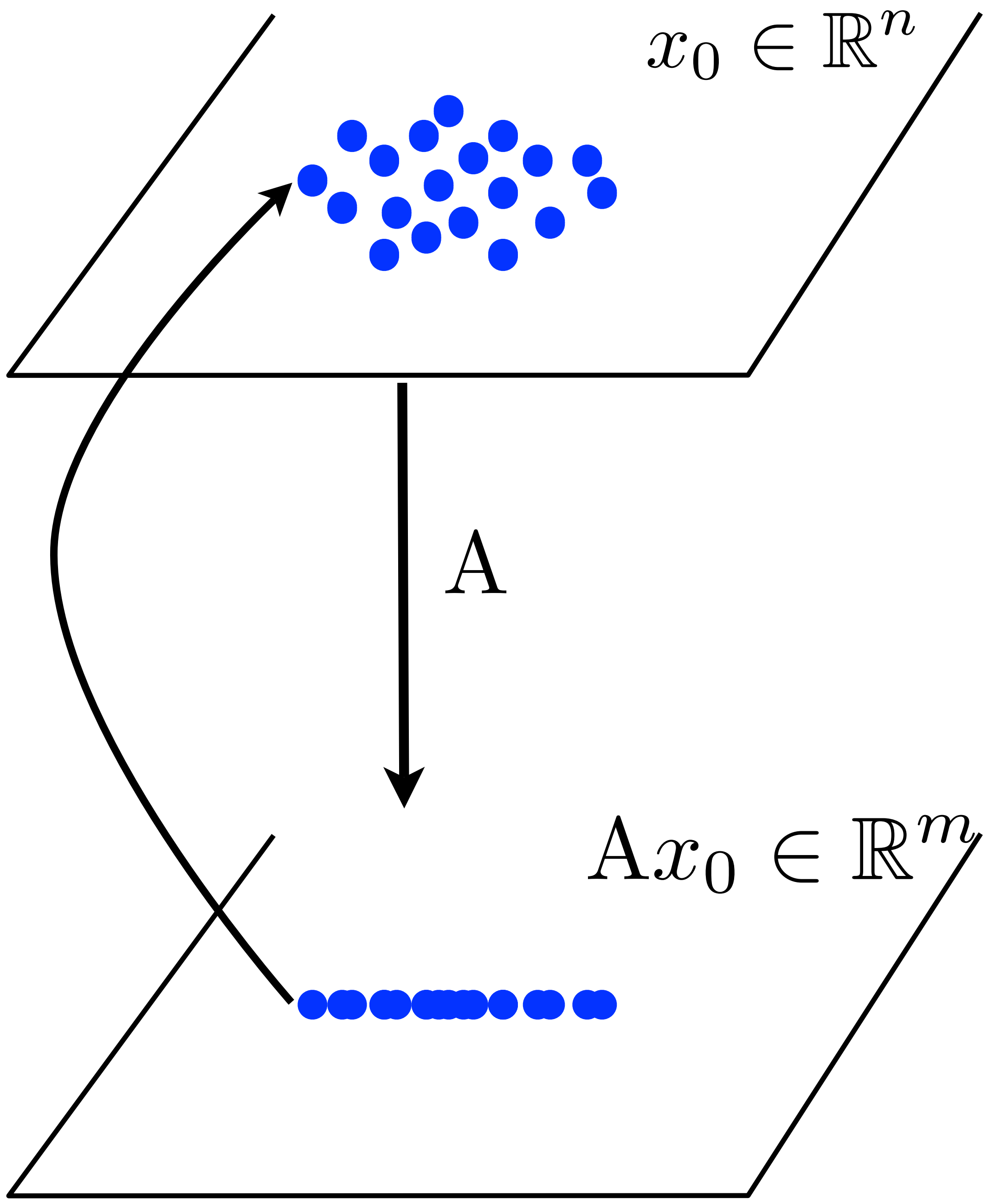
Geometry of inverse problems



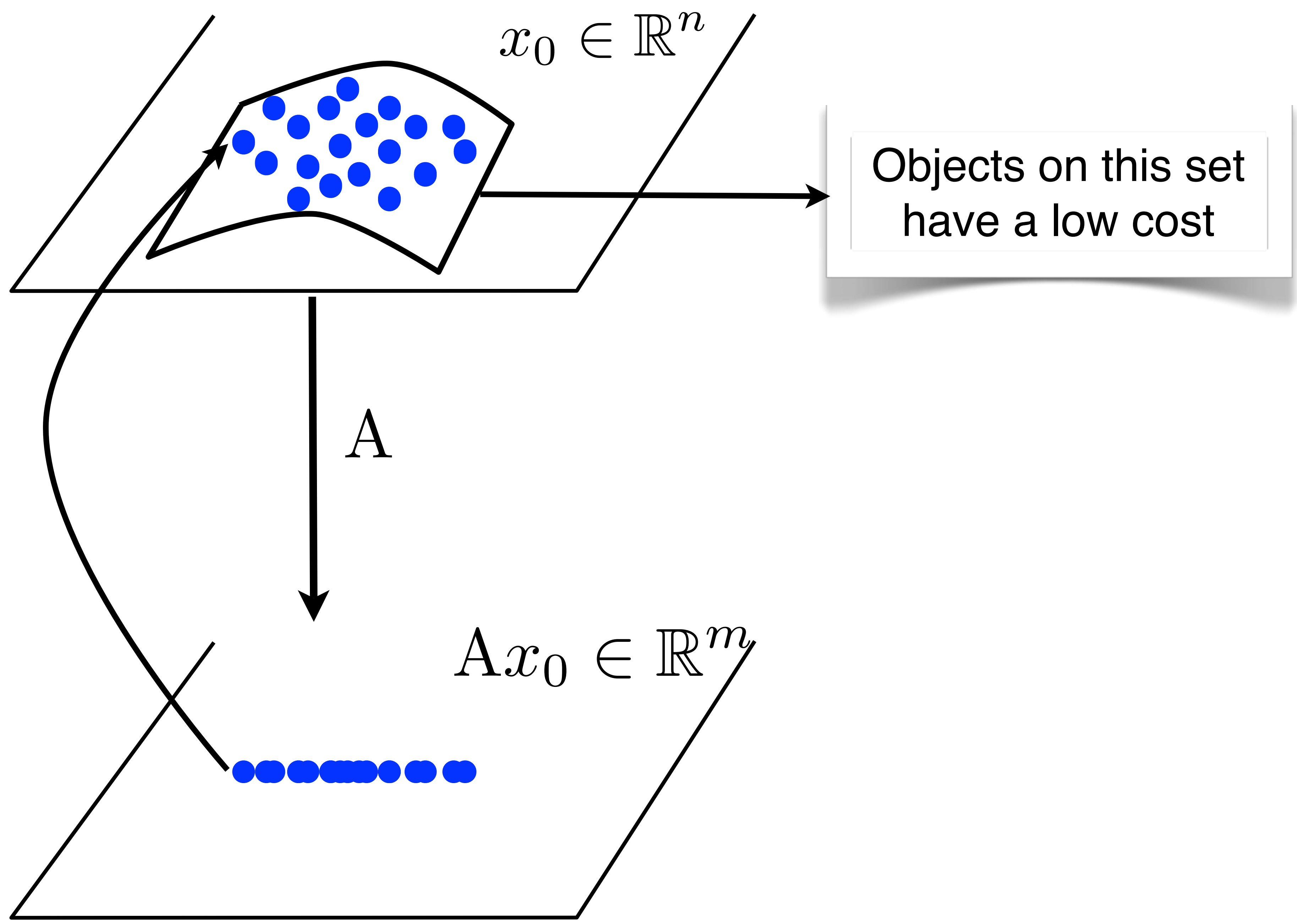
Geometry of inverse problems



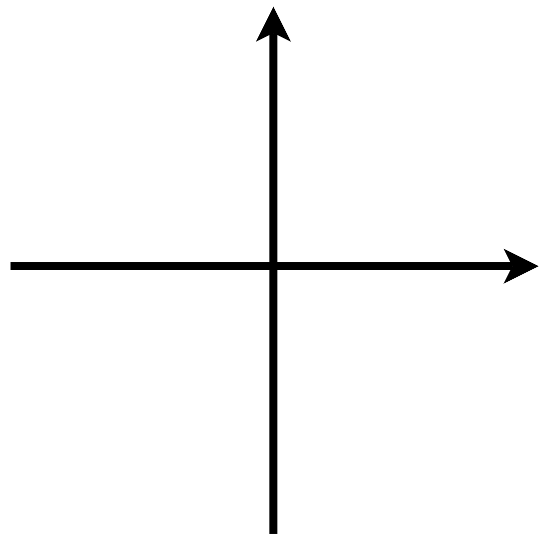
Geometry of inverse problems



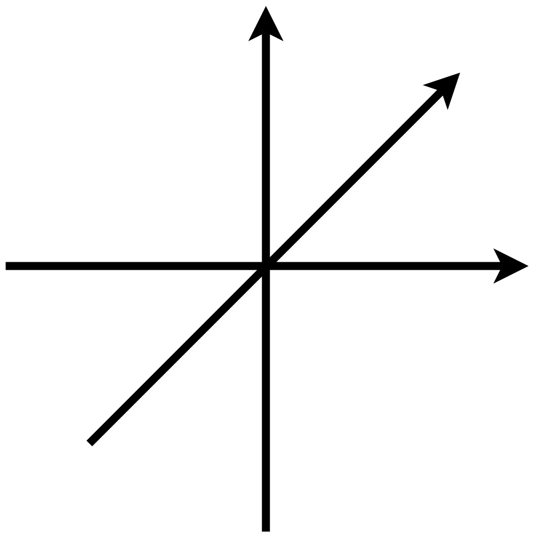
Geometry of inverse problems



Strong notion of sparsity



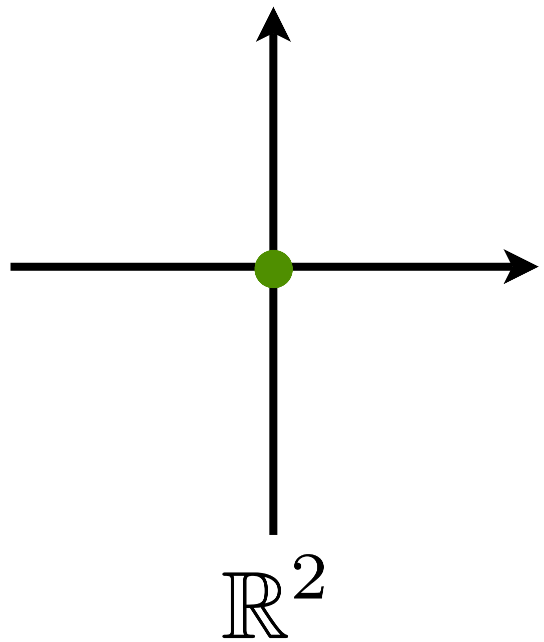
$\mathbb{R}^2$

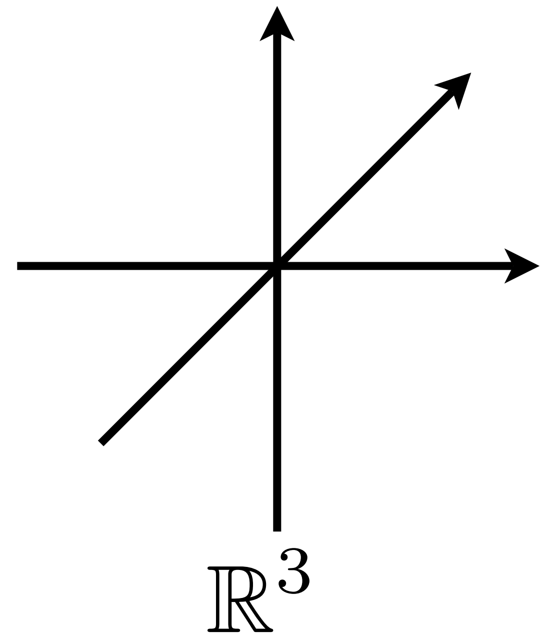


$\mathbb{R}^3$

Strong notion of sparsity

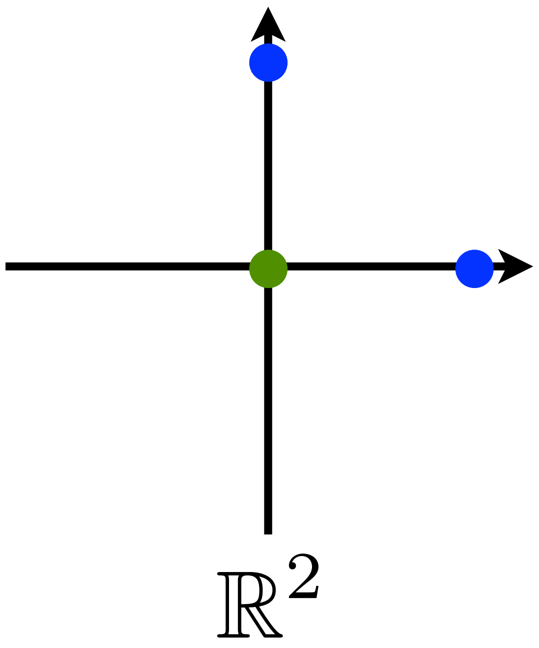
- 0-sparse

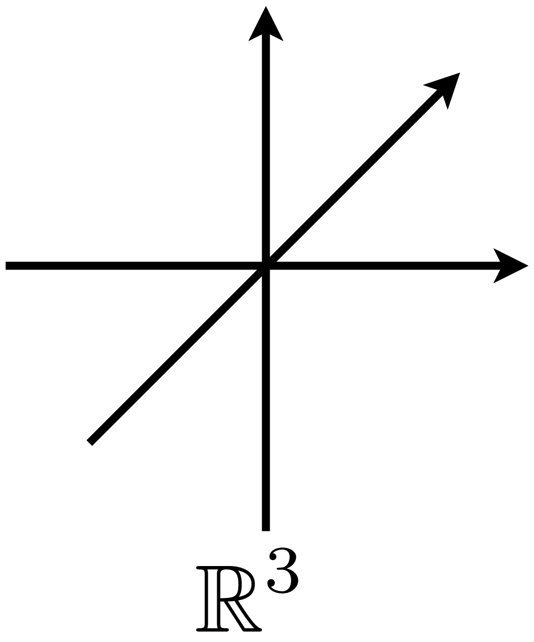




Strong notion of sparsity

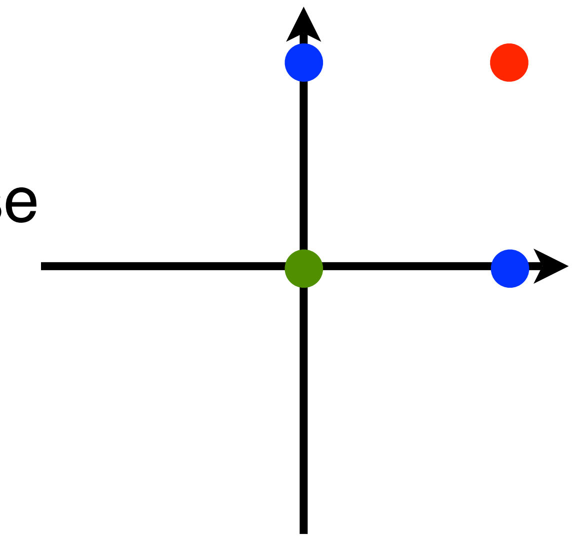
- 0-sparse
- 1-sparse



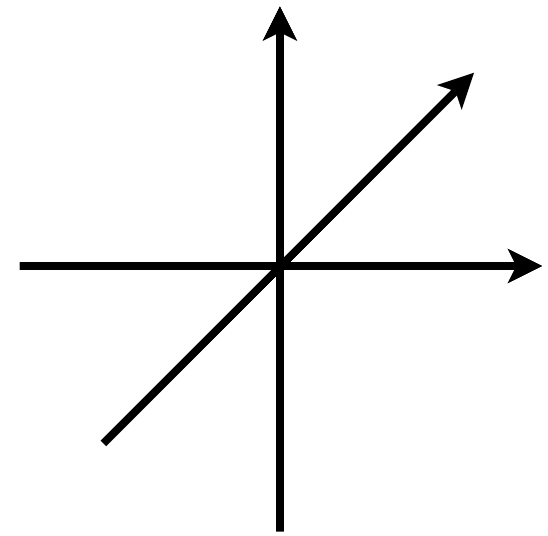


Strong notion of sparsity

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



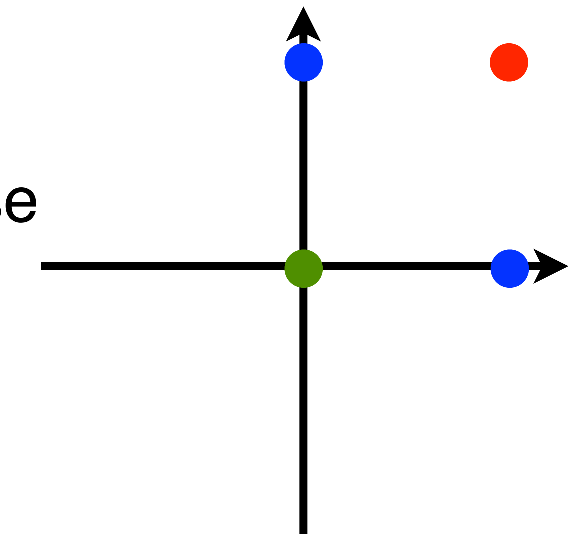
$\mathbb{R}^2$



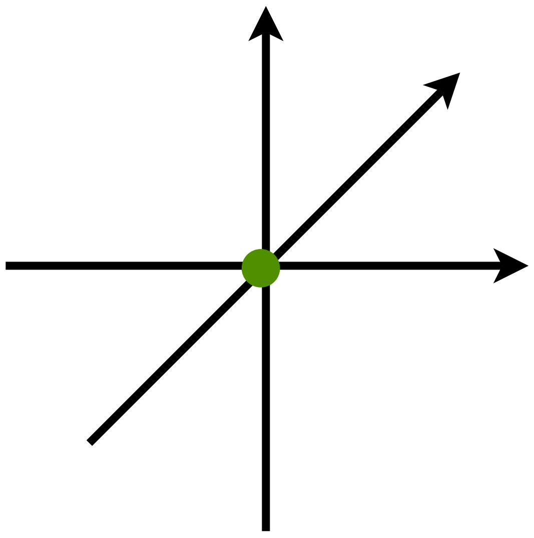
$\mathbb{R}^3$

Strong notion of sparsity

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



$\mathbb{R}^2$

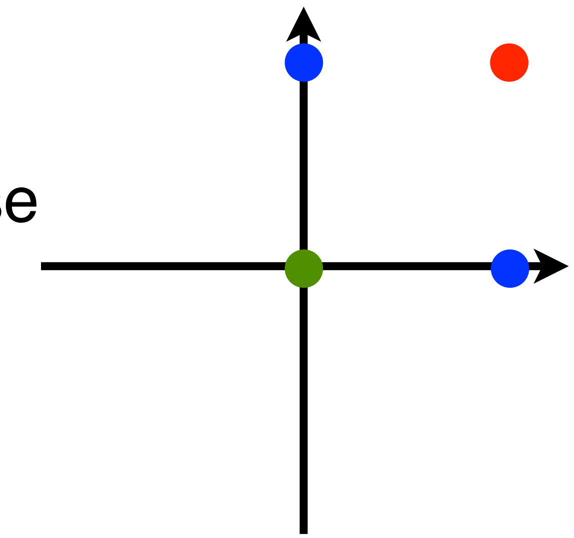


$\mathbb{R}^3$

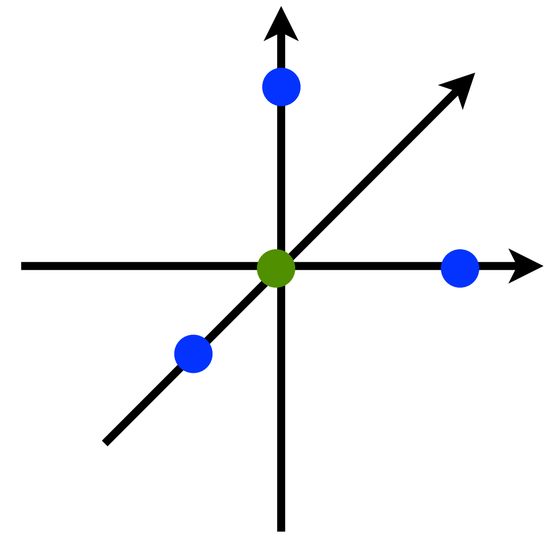
- 0-sparse

Strong notion of sparsity

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



$\mathbb{R}^2$

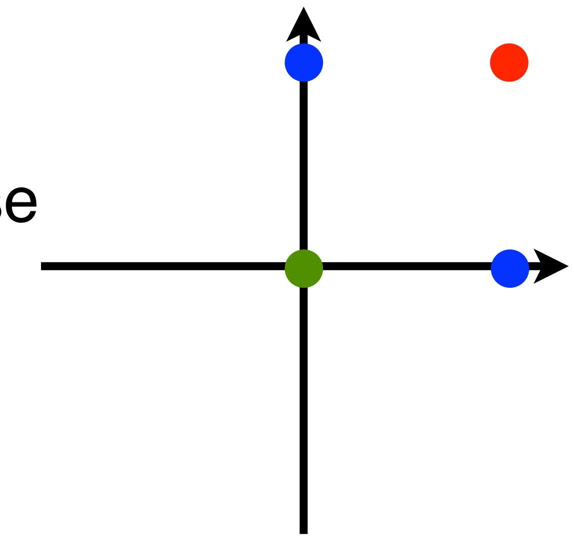


$\mathbb{R}^3$

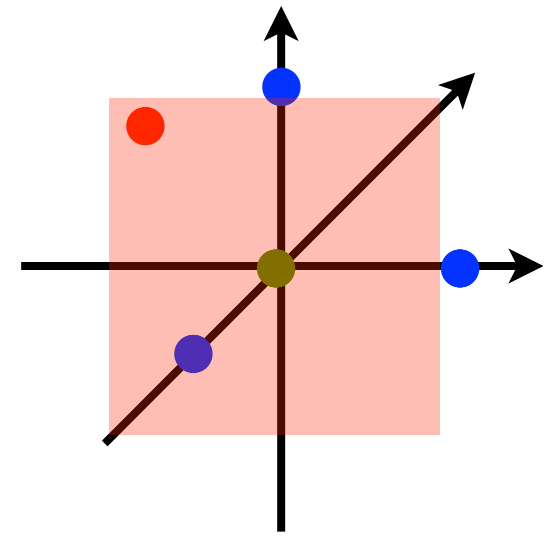
- 0-sparse
- 1-sparse

Strong notion of sparsity

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



$\mathbb{R}^2$

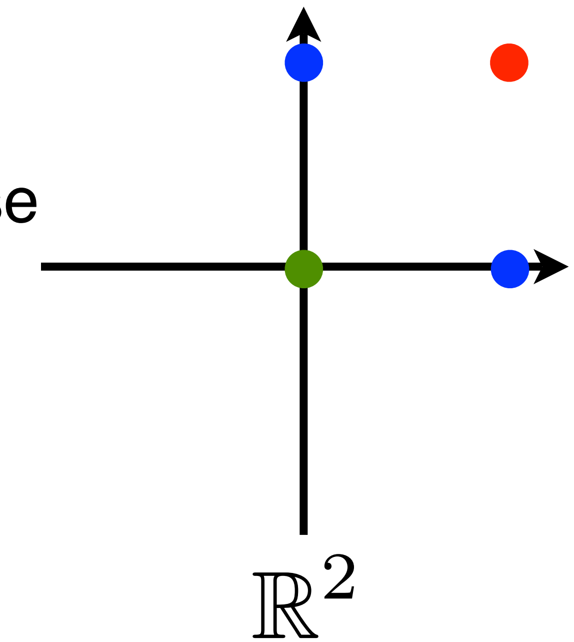


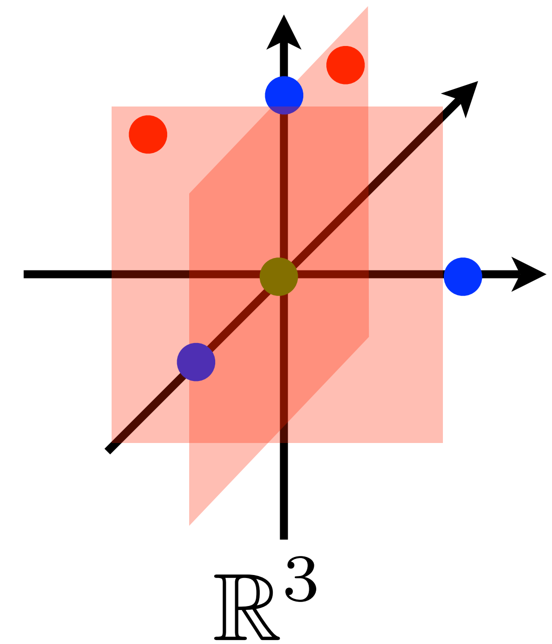
$\mathbb{R}^3$

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

Strong notion of sparsity

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

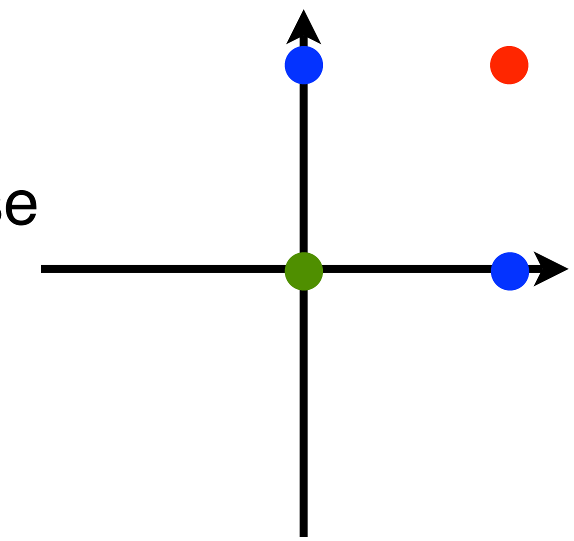




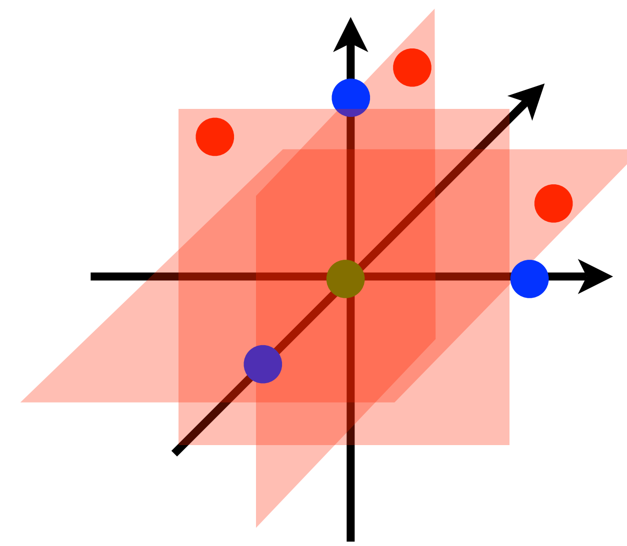
- 0-sparse
- 1-sparse
- 2-sparse

Strong notion of sparsity

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



$\mathbb{R}^2$

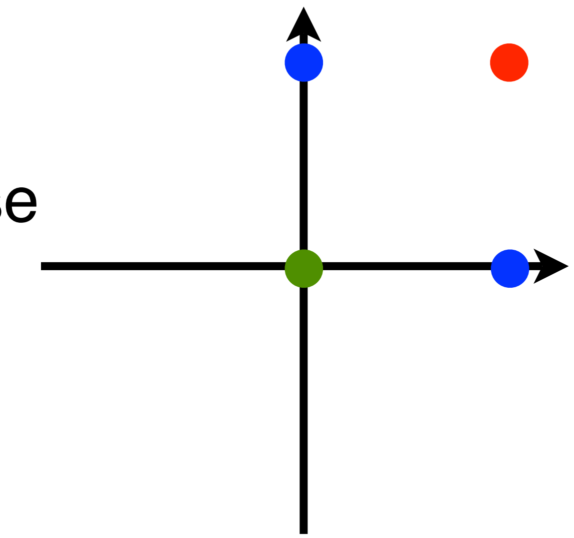


$\mathbb{R}^3$

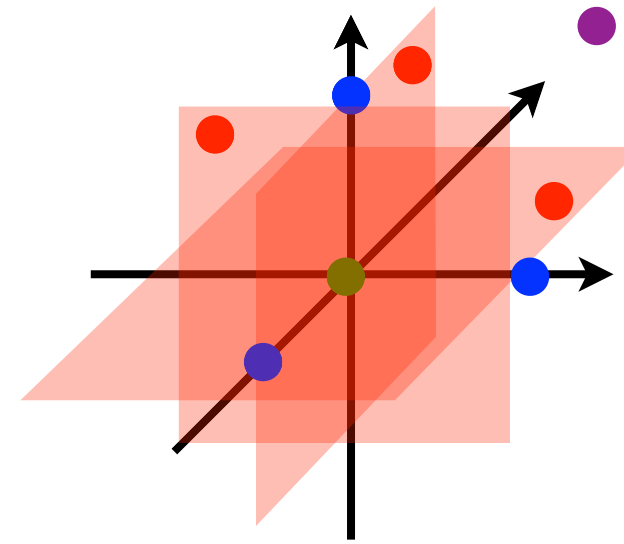
- 0-sparse
- 1-sparse
- 2-sparse

Strong notion of sparsity

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



$\mathbb{R}^2$

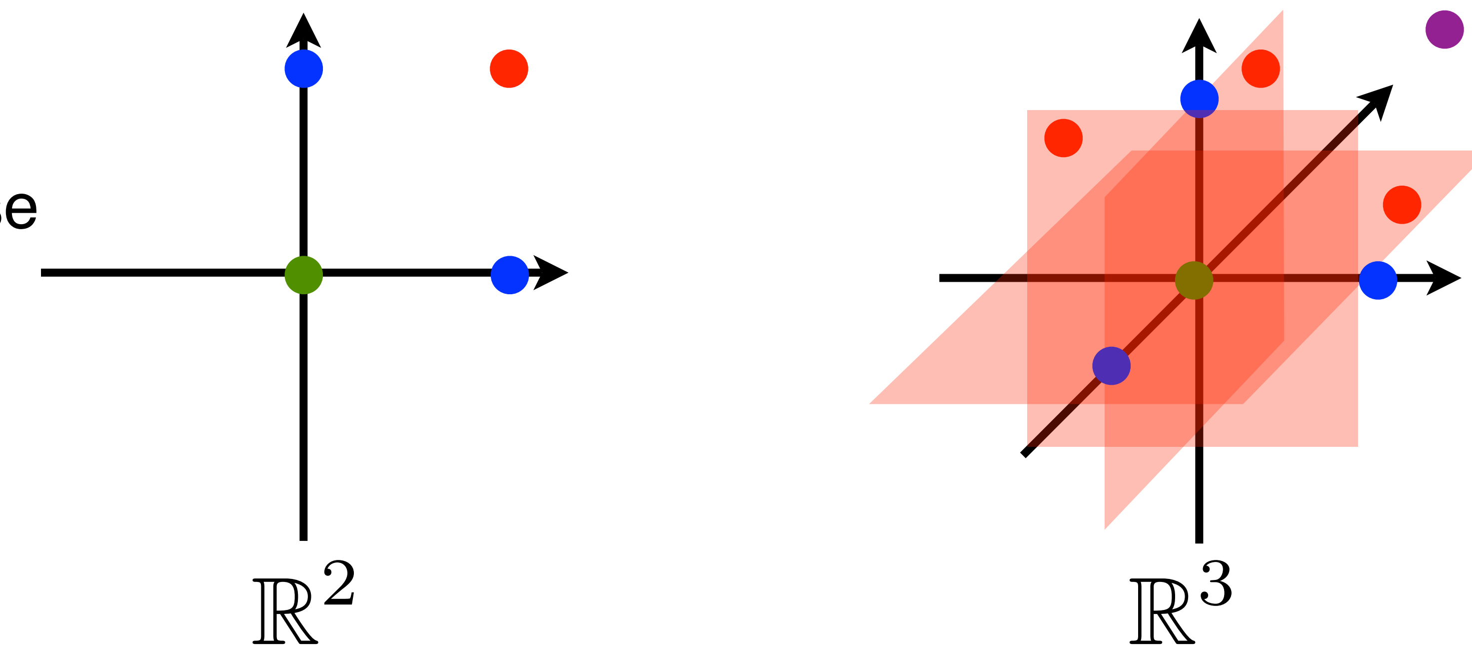


$\mathbb{R}^3$

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

Strong notion of sparsity

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



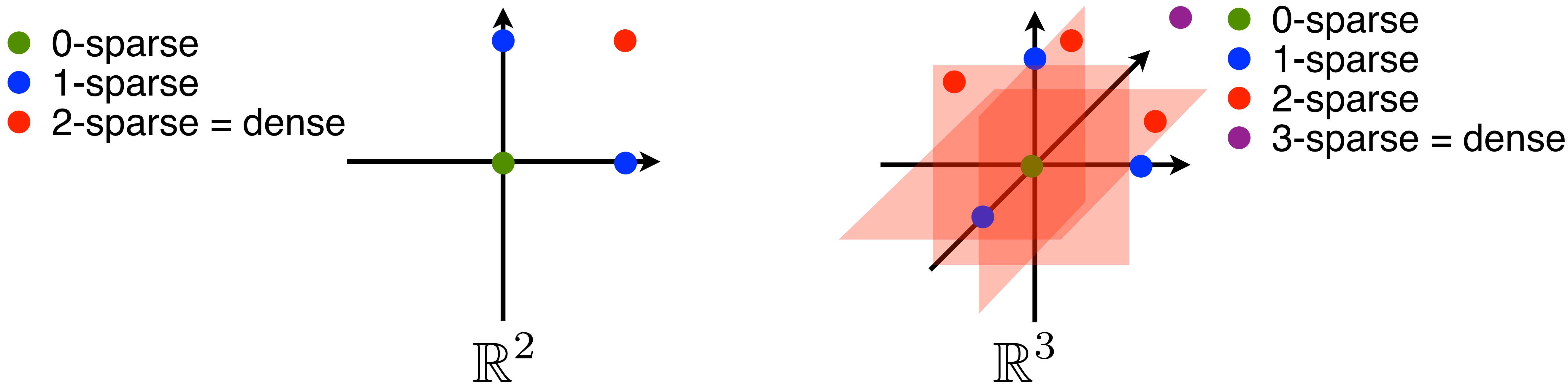
- 0-sparse
- 1-sparse
- 2-sparse
- 3-sparse = dense

$$\operatorname{supp}(x) = \{i = 1, \cdots, n : \; x_i \neq 0\}$$

$$\|x\|_0 = \# \operatorname{supp}(x)$$

(Not a norm : not positively homogenenous)

Strong notion of sparsity



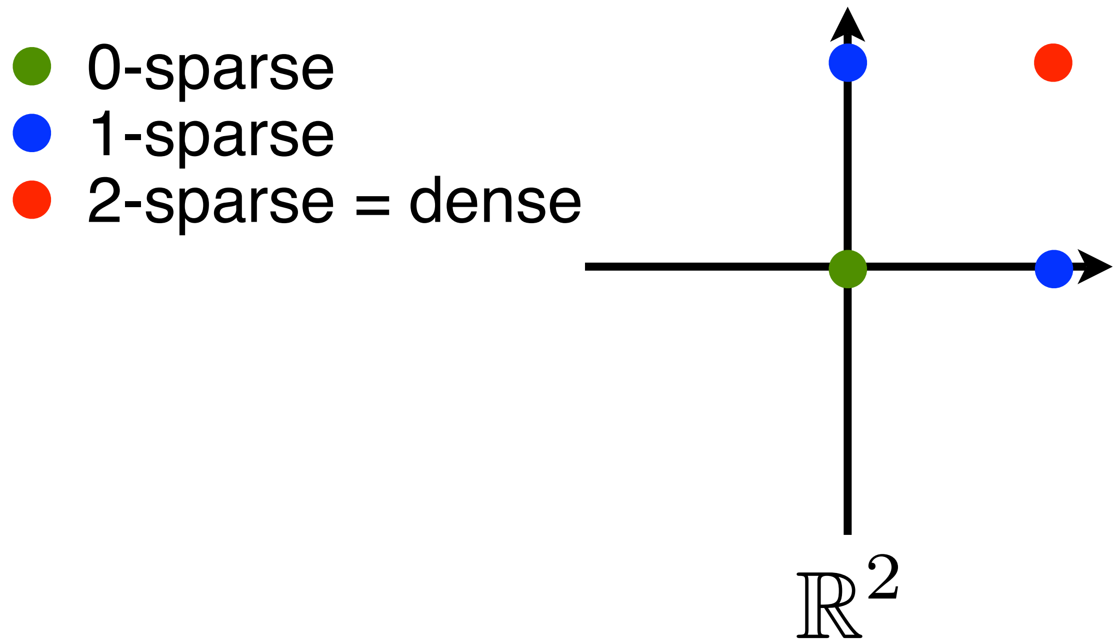
$$\operatorname{supp}(x) = \{i = 1, \cdots, n : \; x_i \neq 0\}$$

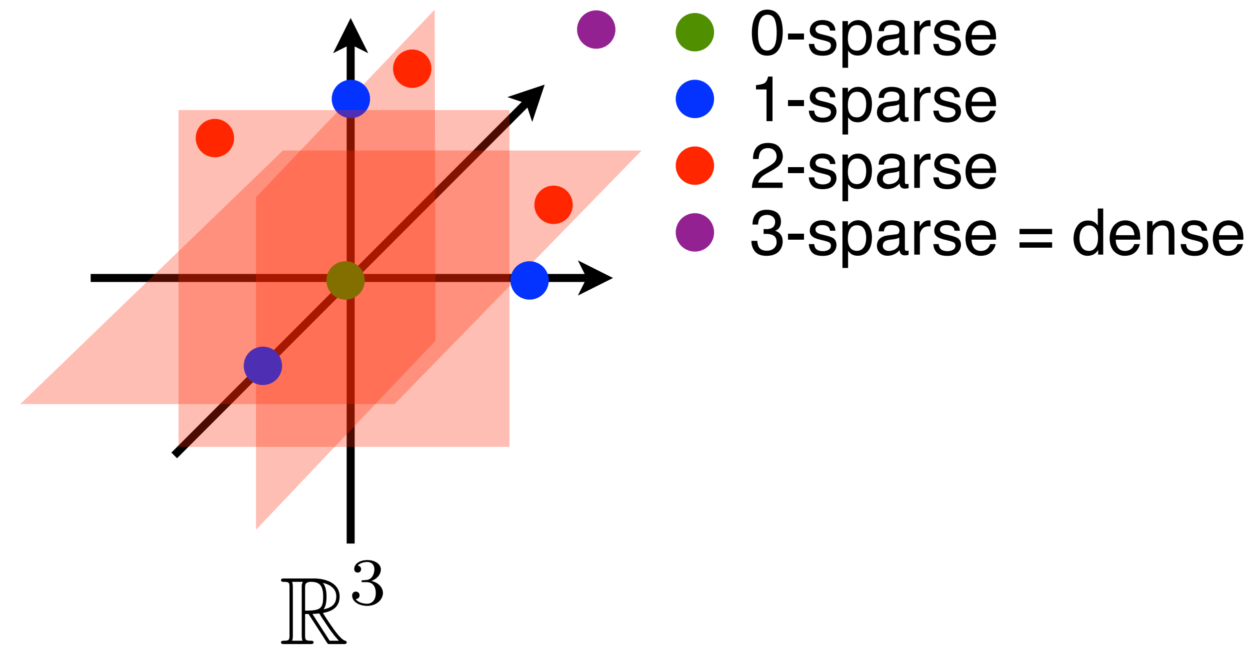
$$\|x\|_0 = \#\operatorname{supp}(x)$$

(Not a norm : not positively homogenenous)

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

Strong notion of sparsity





$$\text{supp}(x) = \{i = 1, \cdots, n : \; x_i \neq 0\}$$

$$\|x\|_0 = \#\,\text{supp}(x)$$

(Not a norm : not positively homogenenous)

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

Model of s-sparse vectors : a union of subspaces

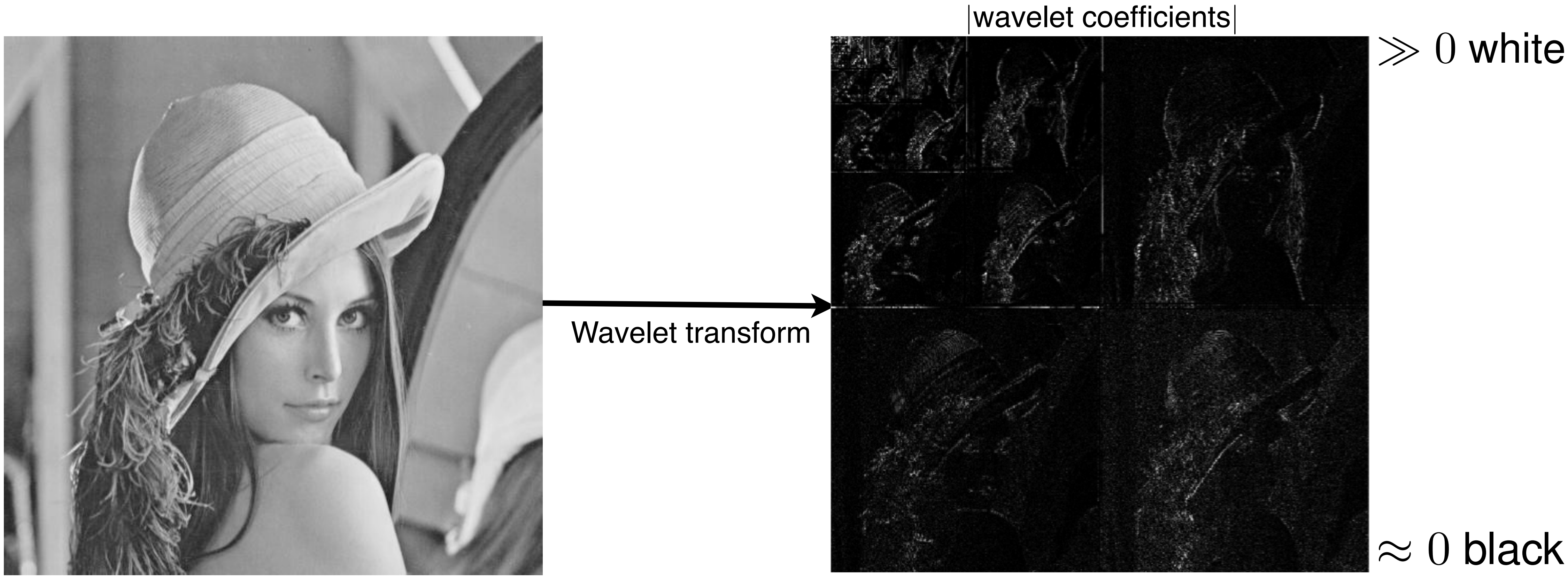
$$\Sigma_s = \bigcup_i \{V_i = \text{span}\left((e_j)_{1 \leq j \leq n}\right) : \; \dim(V_i) = s\}.$$

Weak notion of sparsity

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

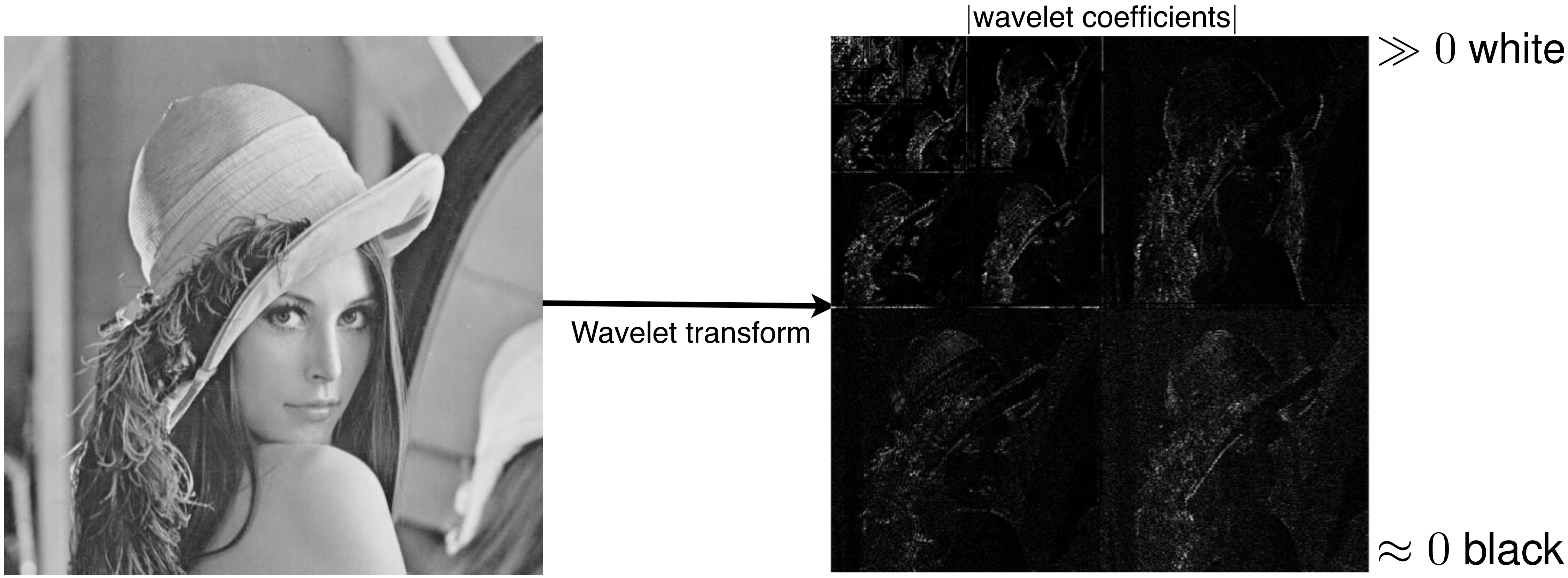
Weak notion of sparsity

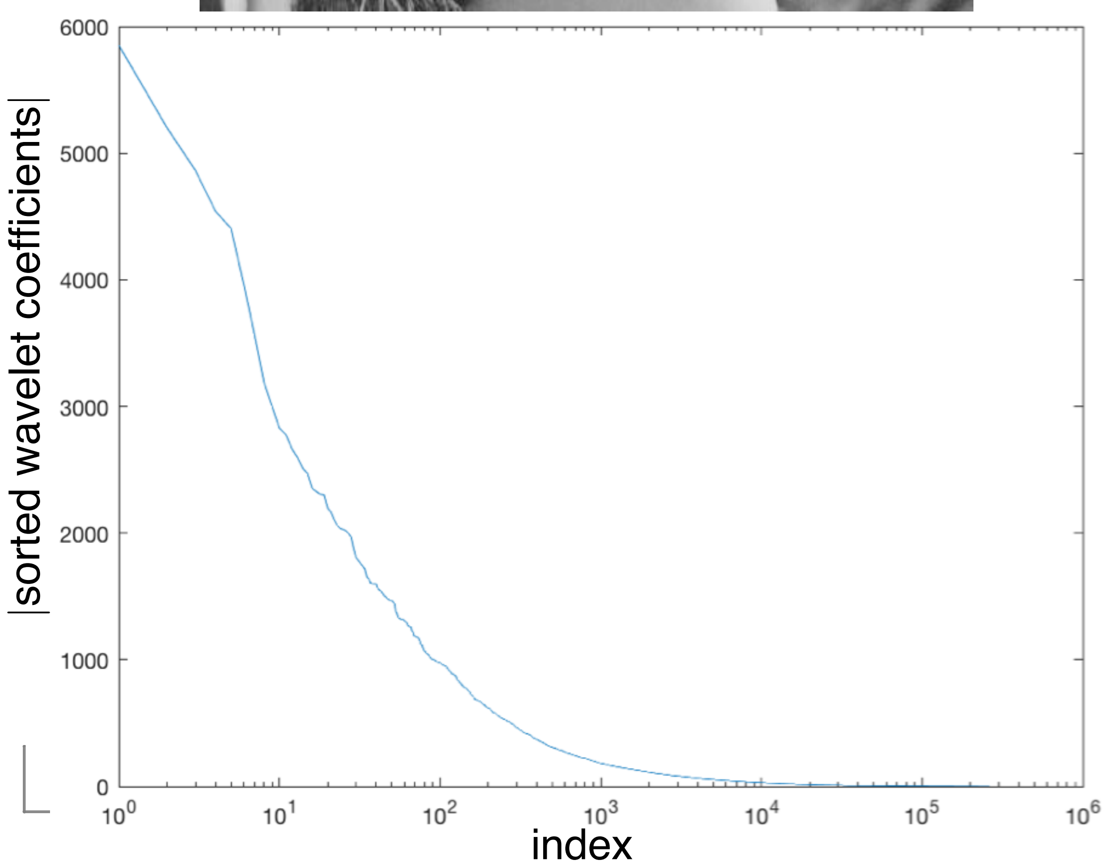
- In nature, signals, images, information, are not (strongly) sparse.



Weak notion of sparsity

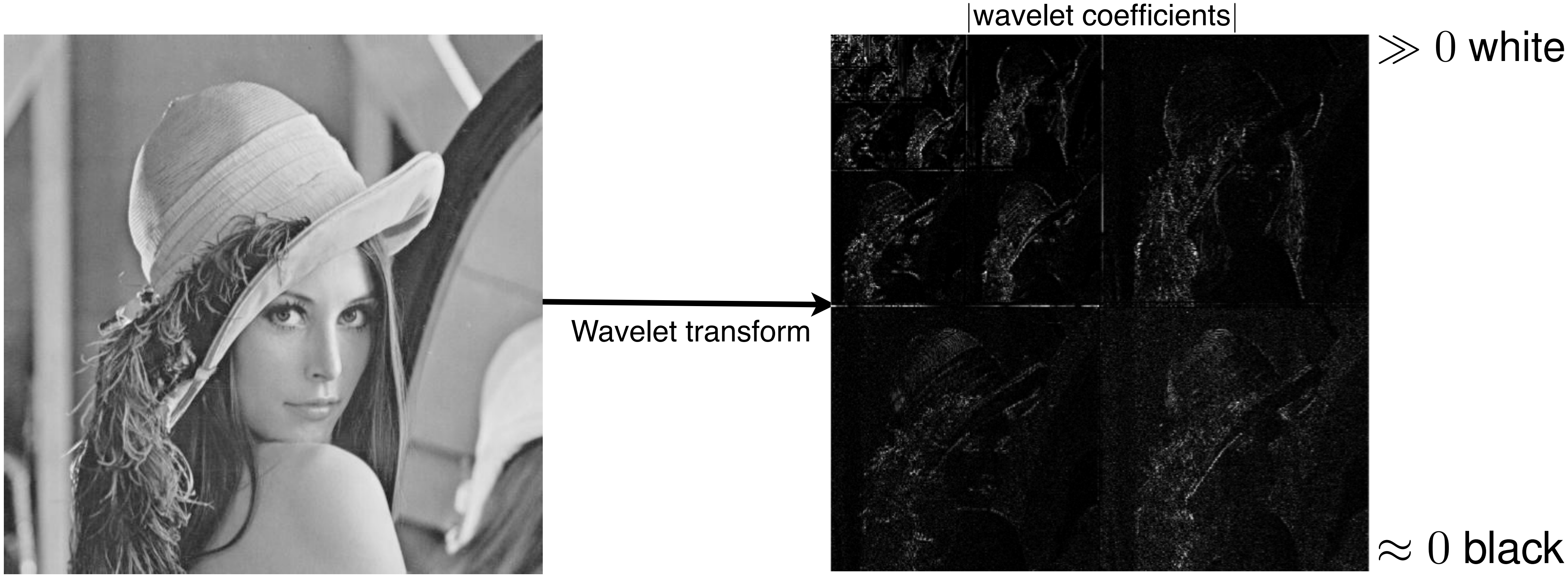
- In nature, signals, images, information, are not (strongly) sparse.

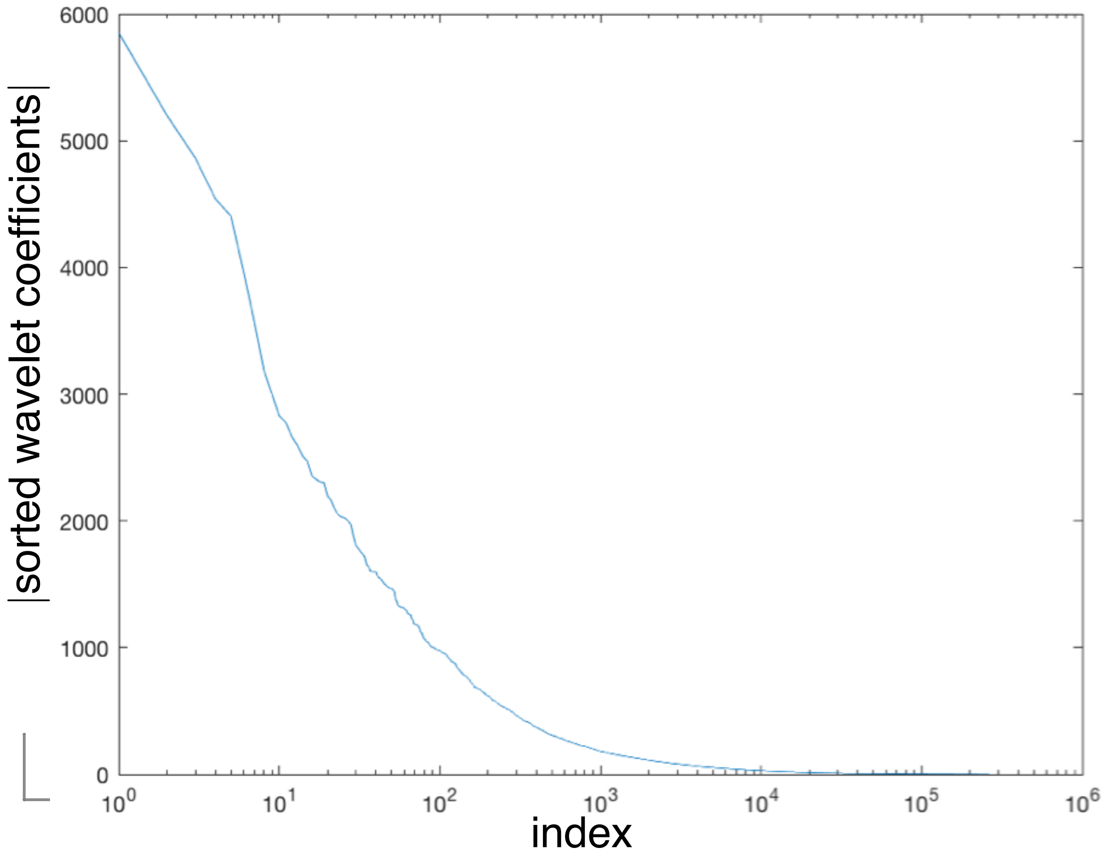




Weak notion of sparsity

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





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 $y = \mathrm{A}x$ where x is sparse

- If $\|x\|_0 \leq m$ and A_I is full-rank ($I \stackrel{\mathrm{def}}{=} \mathrm{supp}(x)$), 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 $y = \mathrm{A}x$ where x is sparse

Regularization

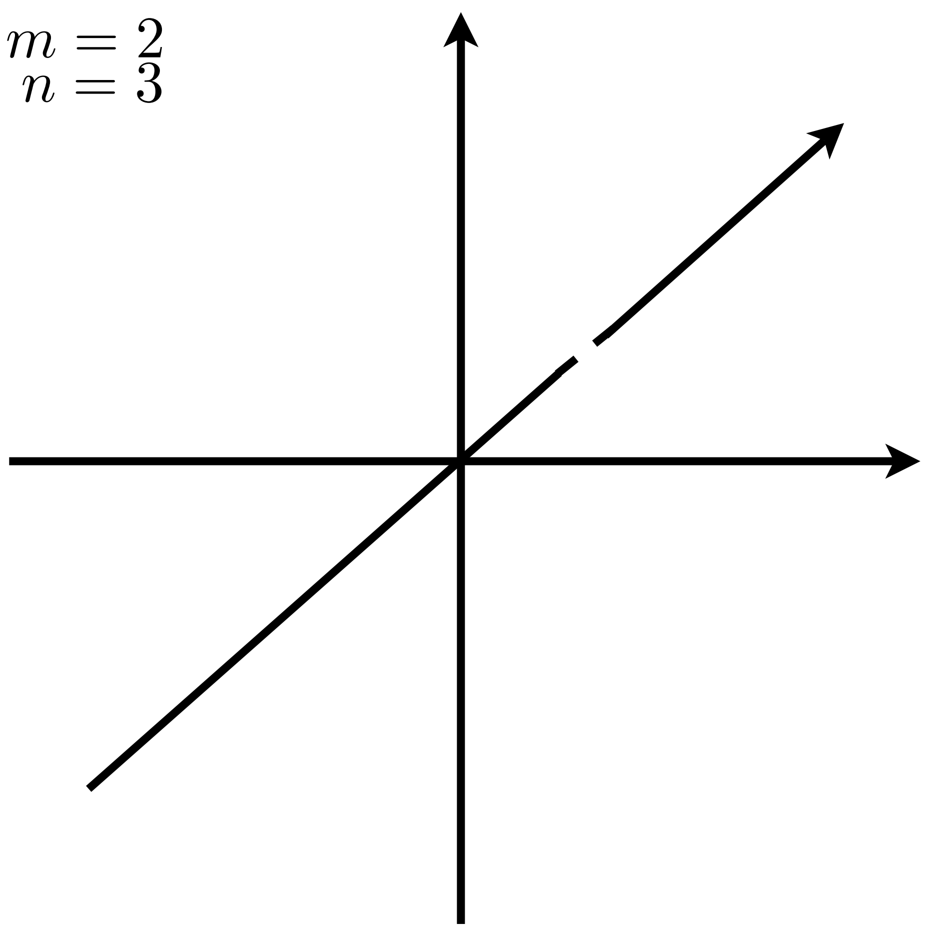
Solve $y = \mathrm{A}x$ where x is sparse

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

Regularization

Solve $y = \mathrm{A}x$ where x is sparse

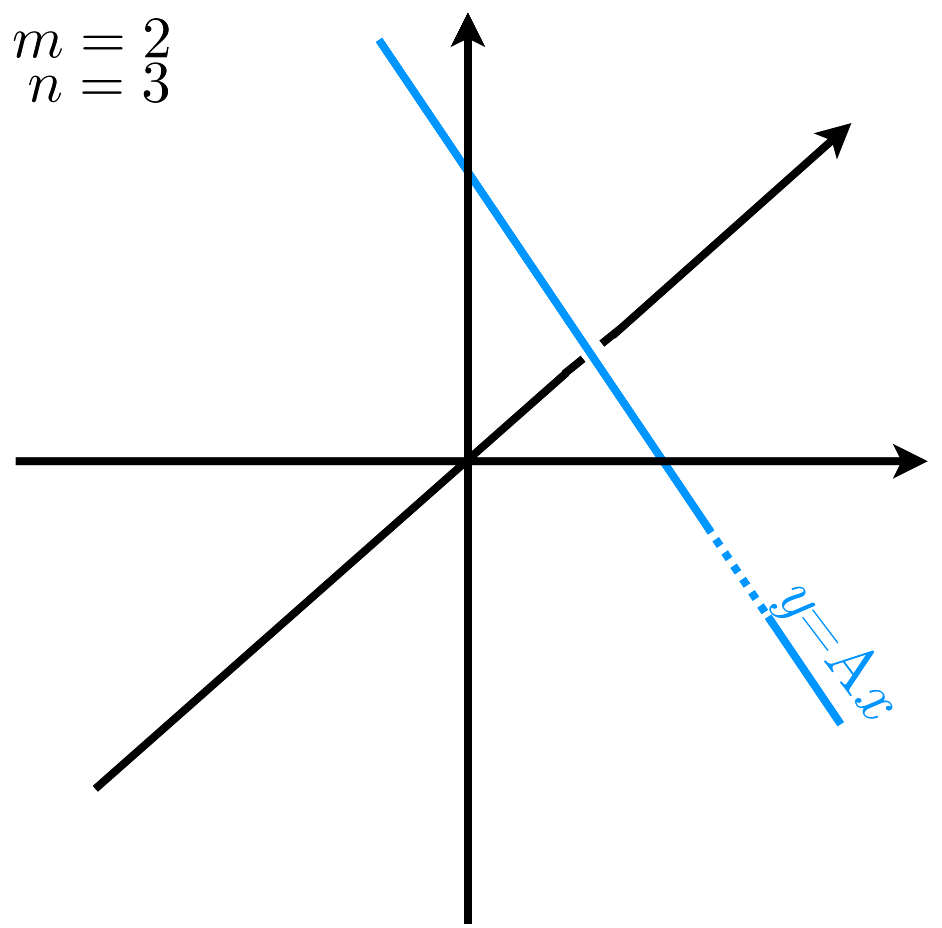
$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{such that} \quad y = \mathrm{A}x$$



Regularization

Solve $y = \mathrm{A}x$ where x is sparse

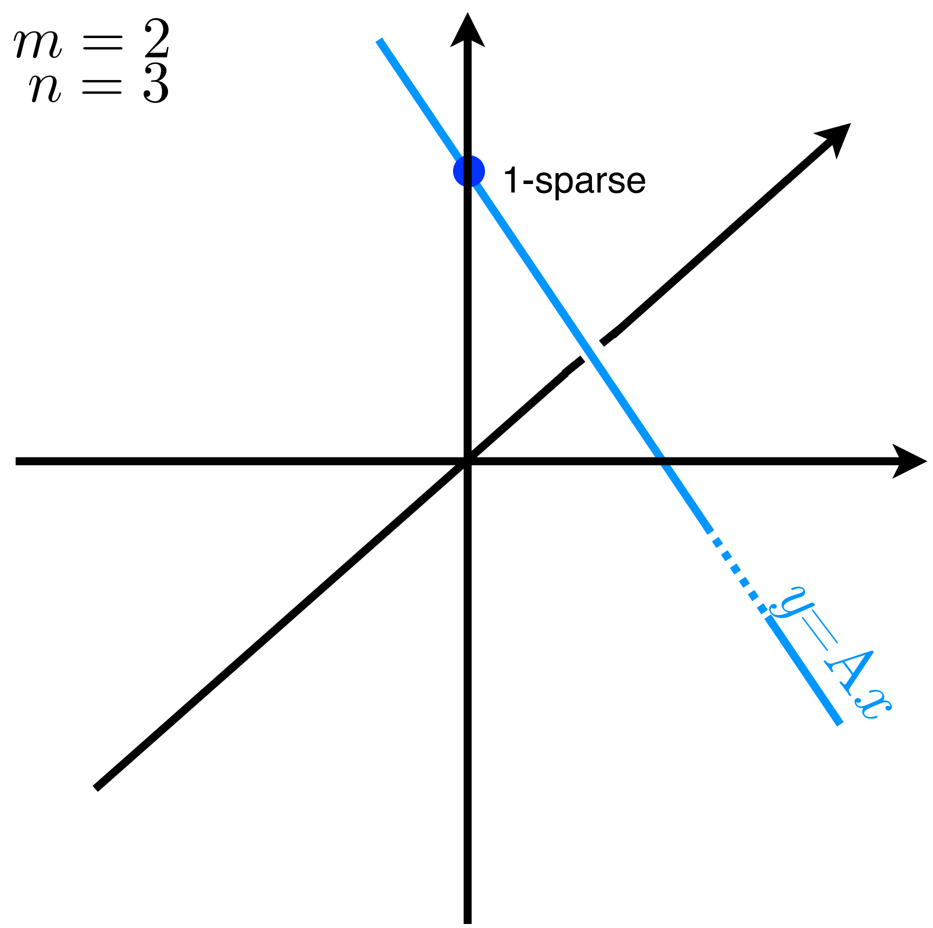
$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{such that} \quad y = \mathrm{A}x$$



Regularization

Solve $y = \mathrm{A}x$ where x is sparse

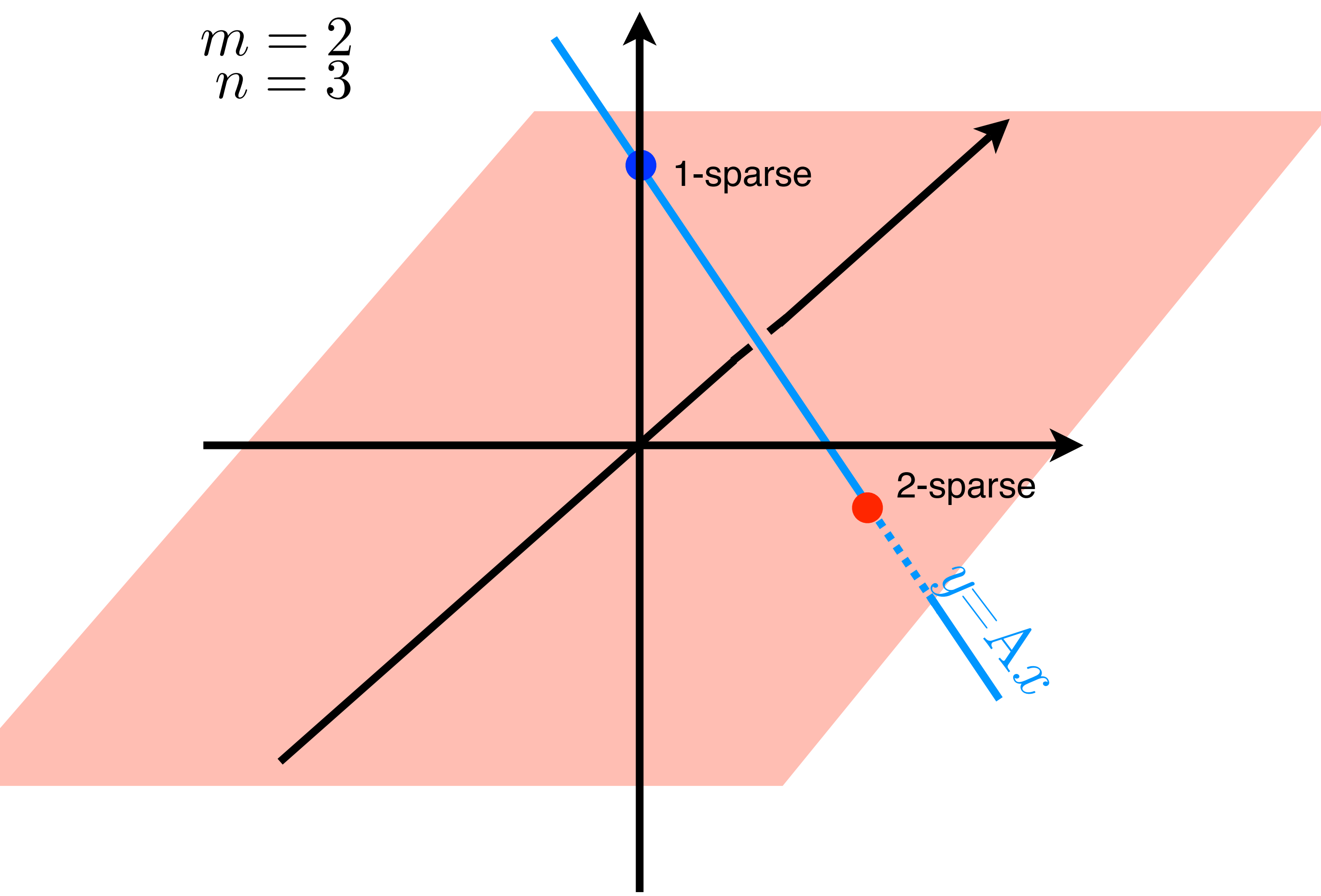
$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{such that} \quad y = \mathrm{A}x$$



Regularization

Solve $y = \mathrm{A}x$ where x is sparse

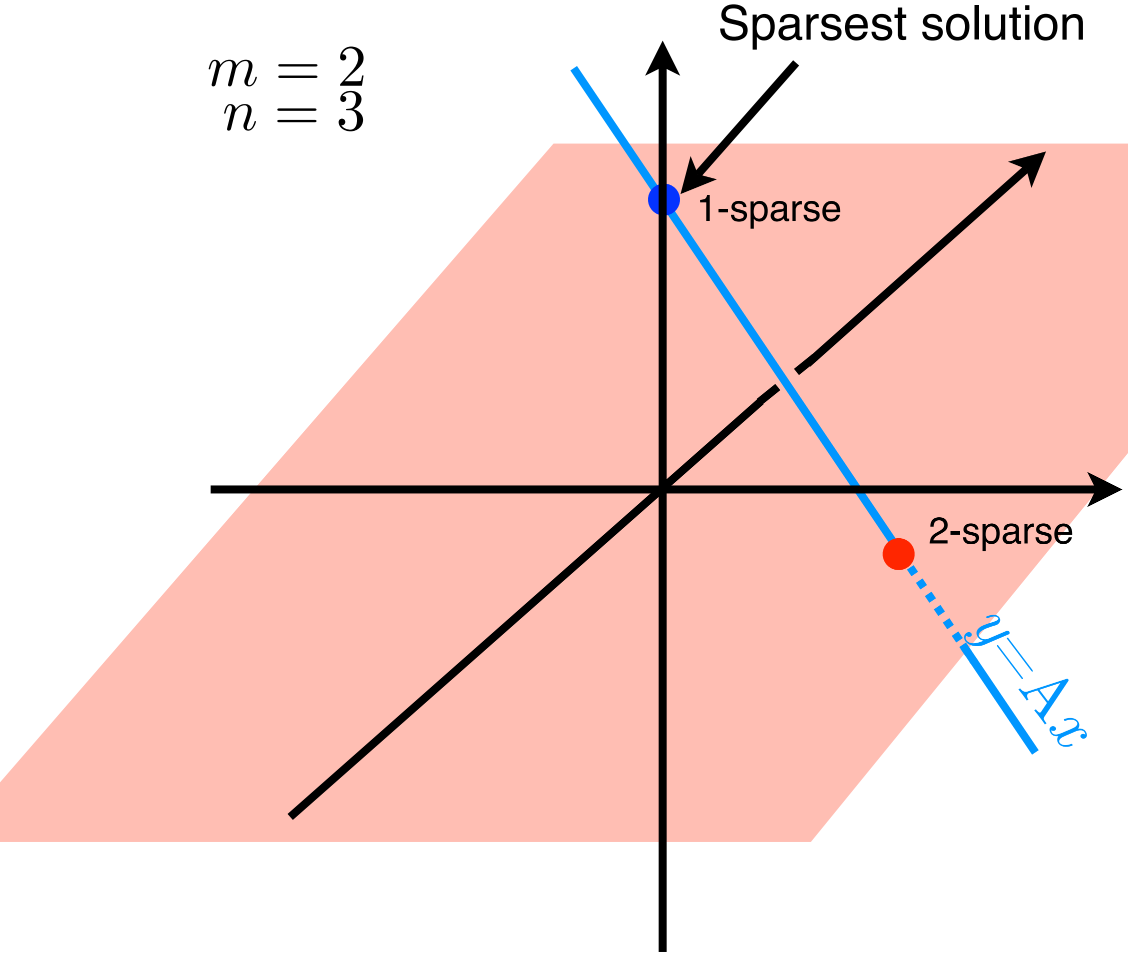
$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{such that} \quad y = \mathrm{A}x$$



Regularization

Solve $y = \mathrm{A}x$ where x is sparse

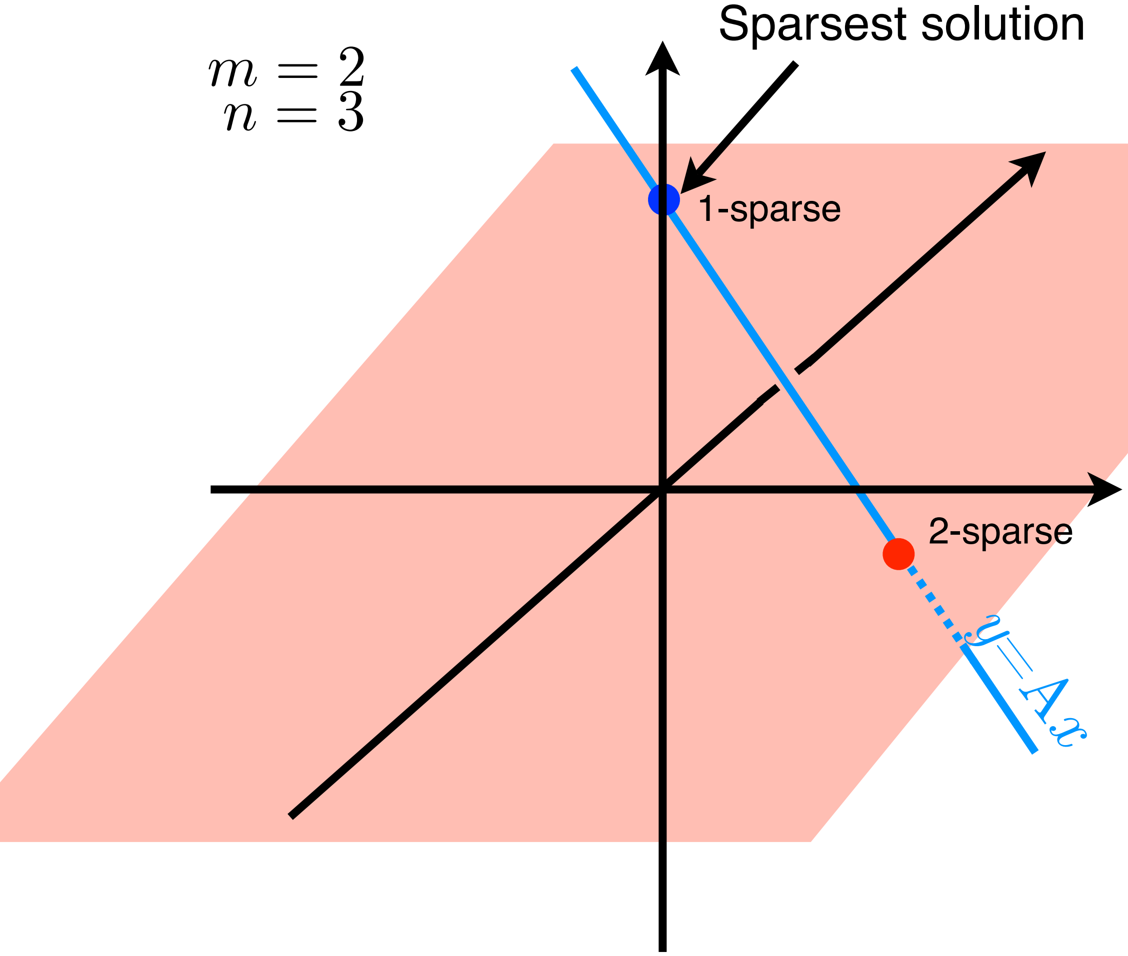
$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{such that} \quad y = \mathrm{A}x$$



Regularization

Solve $y = \mathrm{A}x$ where x is sparse

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



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

Relaxation

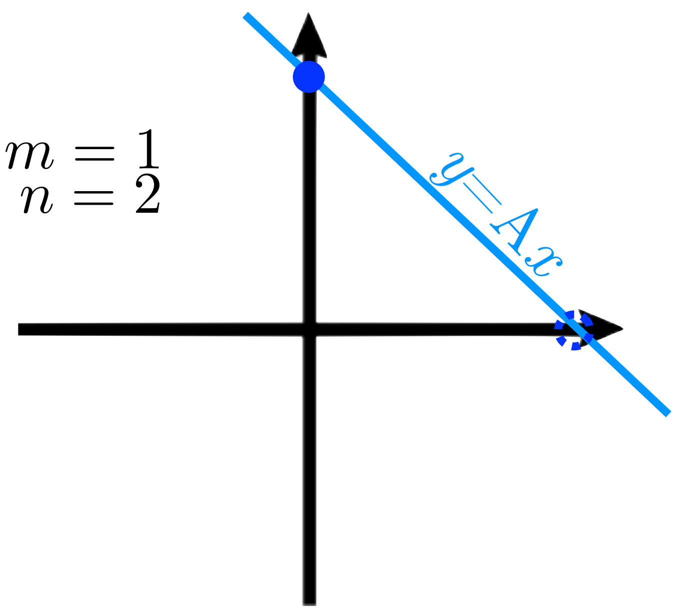
Solve $y = \mathrm{A}x$ where x is sparse

Relaxation

Solve $y = \mathrm{A}x$ where x is sparse

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

$m = 1$
$n = 2$

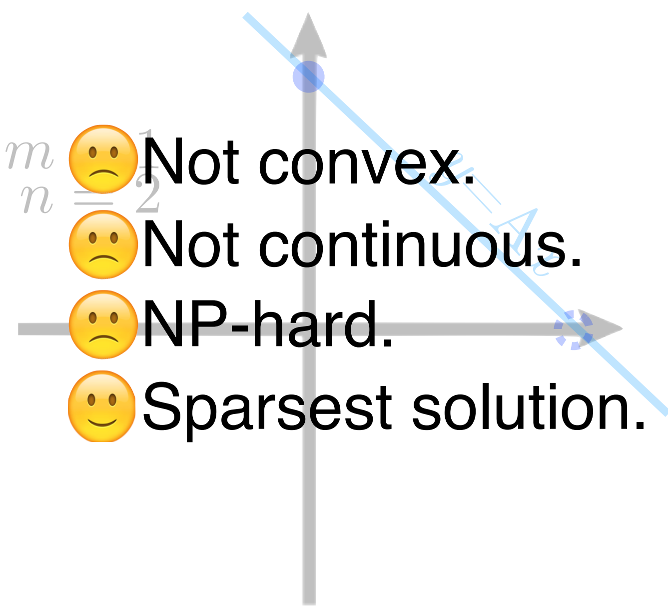


Relaxation

Solve $y = Ax$ where x is sparse

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

$m = 1$
$n = 2$



- 🙁 Not convex.
- 🙁 Not continuous.
- 🙁 NP-hard.
- 🙂 Sparsest solution.

Relaxation

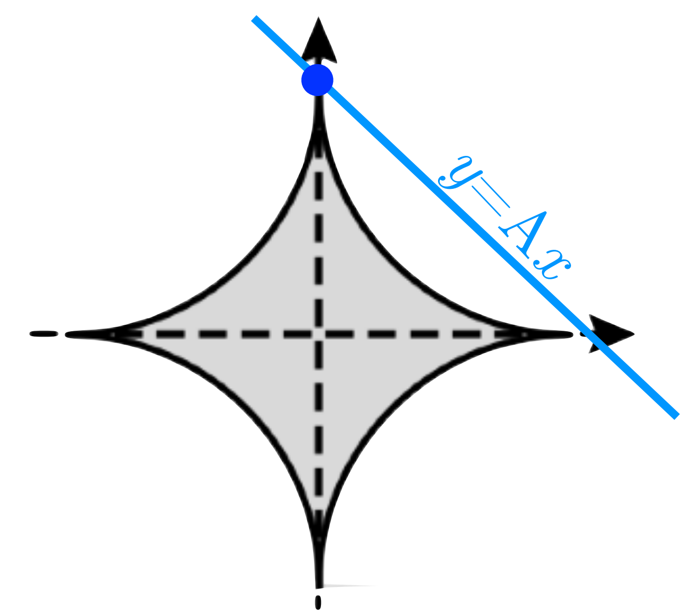
Solve $y = \mathrm{A}x$ where x is sparse

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

$m = 1$
$n = 2$

- 🙁 Not convex.
- 🙁 Not continuous.
- 🙁 NP-hard.
- 🙂 Sparsest solution.

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



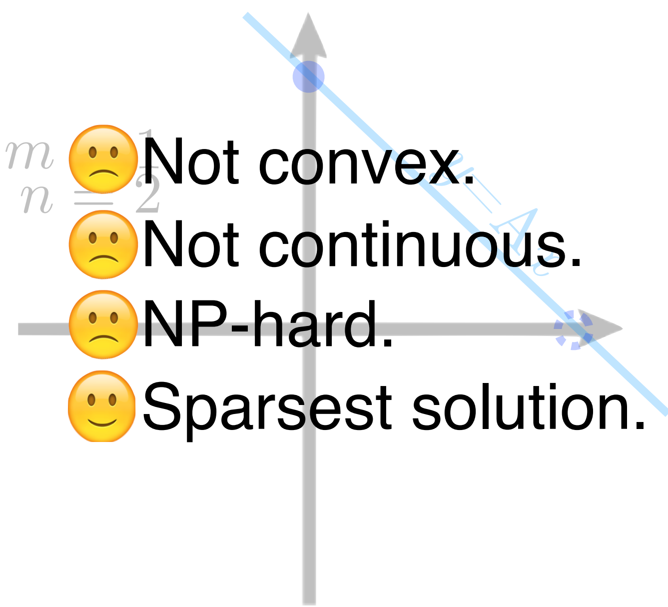
$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

Relaxation

Solve $y = Ax$ where x is sparse

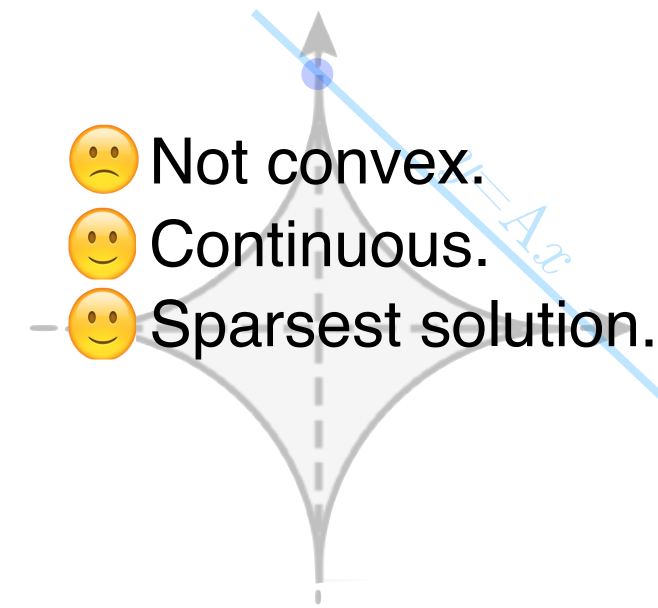
$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{s.t.} \quad y = Ax$$

$m = 1$
$n = 2$



- 🙁 Not convex.
- 🙁 Not continuous.
- 🙁 NP-hard.
- 🙂 Sparsest solution.

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



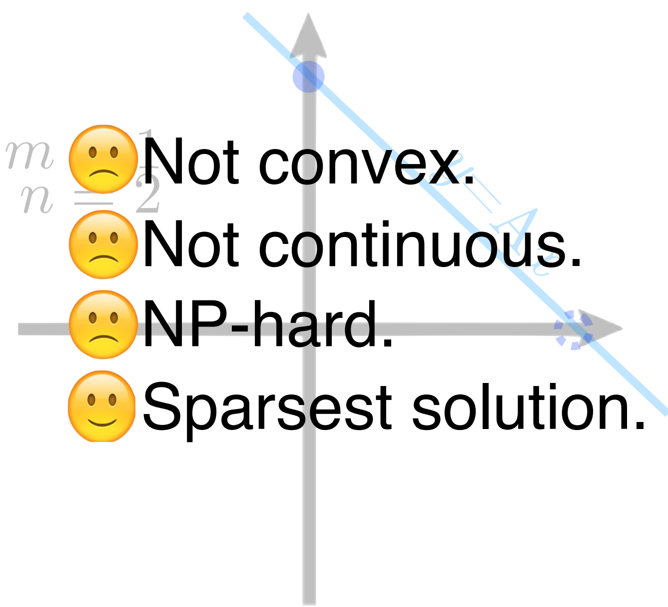
- 🙁 Not convex.
- 🙂 Continuous.
- 🙂 Sparsest solution.

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

Relaxation

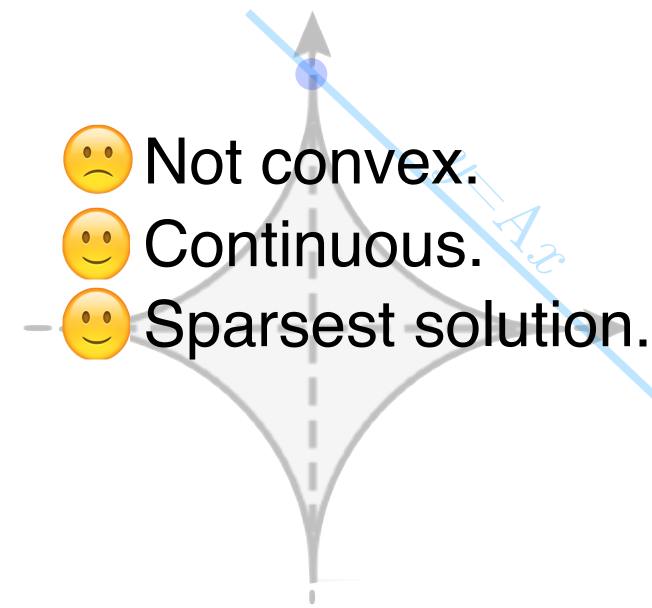
Solve $y = \mathrm{A}x$ where x is sparse

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



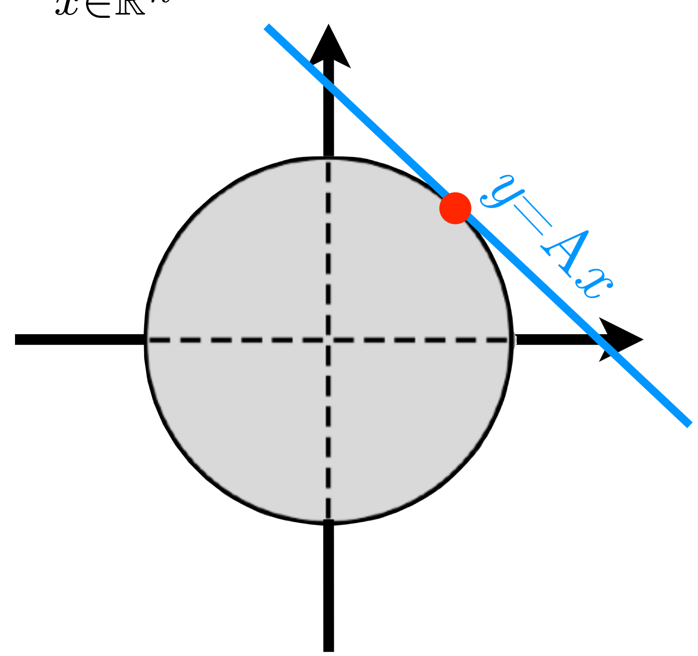
- 🙁 Not convex.
- 🙁 Not continuous.
- 🙁 NP-hard.
- 🙂 Sparsest solution.

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



- 🙁 Not convex.
- 🙂 Continuous.
- 🙂 Sparsest solution.

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

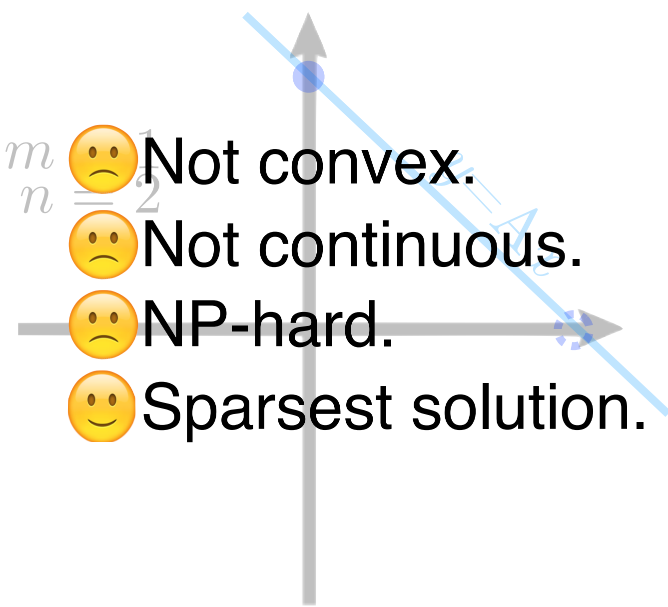


$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

Relaxation

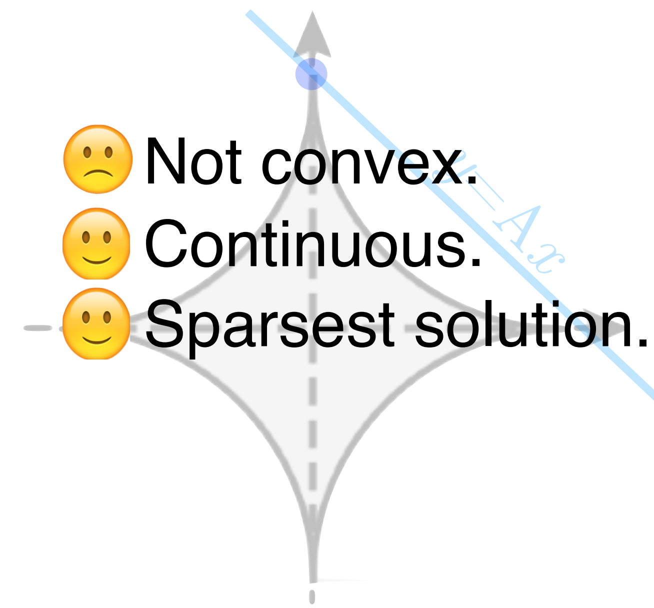
Solve $y = Ax$ where x is sparse

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



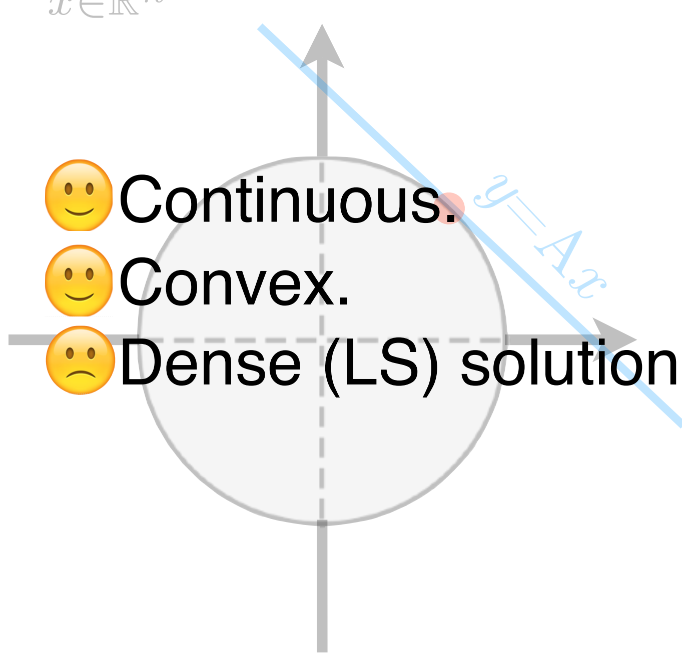
- 🙁 Not convex.
- 🙁 Not continuous.
- 🙁 NP-hard.
- 🙂 Sparsest solution.

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



- 🙁 Not convex.
- 🙂 Continuous.
- 🙂 Sparsest solution.

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



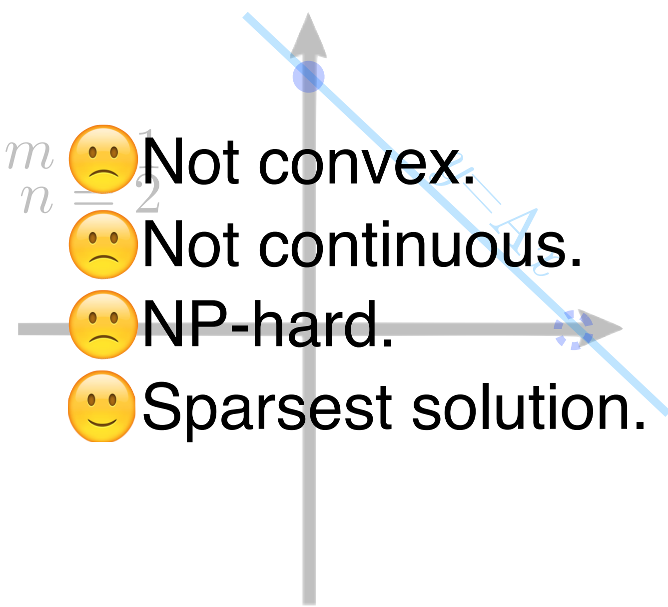
- 🙂 Continuous.
- 🙂 Convex.
- 🙁 Dense (LS) solution

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

Relaxation

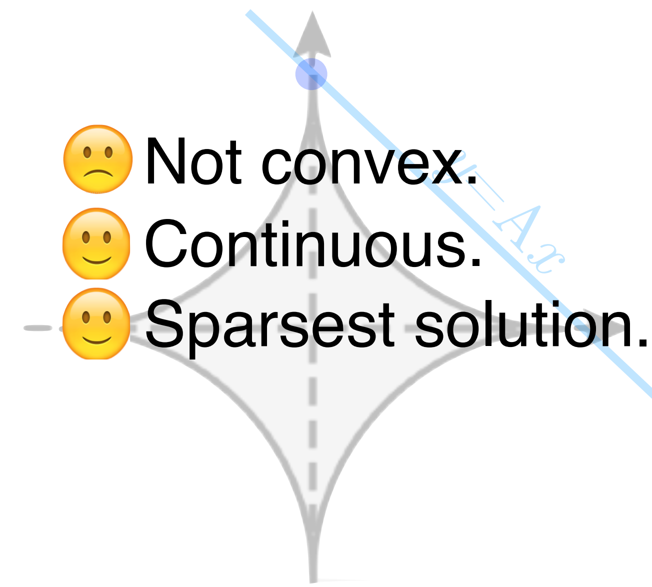
Solve $y = \mathrm{A}x$ where x is sparse

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



- Not convex.
- Not continuous.
- NP-hard.
- Sparsest solution.

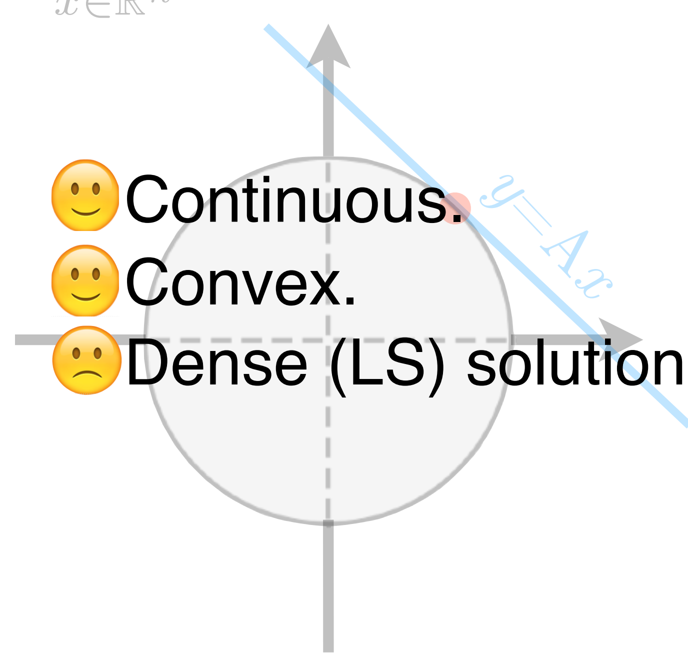
$\min_{x \in \mathbb{R}^n} \|x\|_{0.5} \quad$ s.t. $\quad y = \mathrm{A}x$



- Not convex.
- Continuous.
- Sparsest solution.

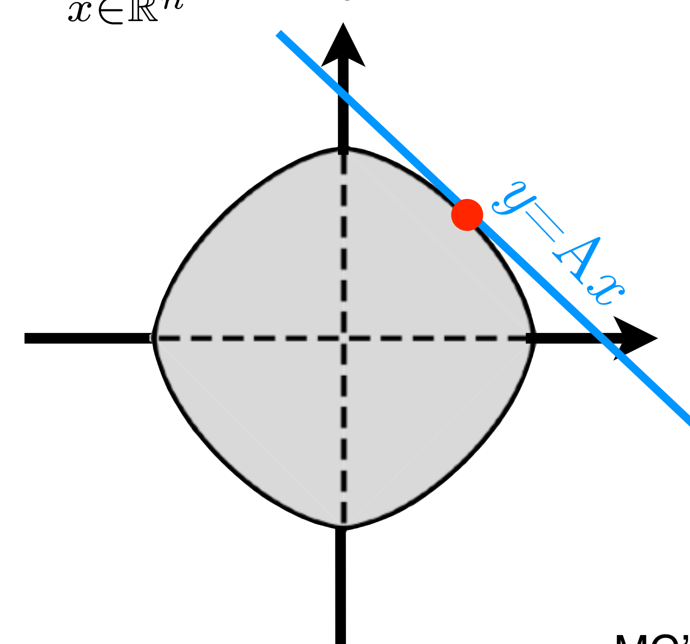
$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

$\min_{x \in \mathbb{R}^n} \|x\|_2 \quad$ s.t. $\quad y = \mathrm{A}x$



- Continuous.
- Convex.
- Dense (LS) solution

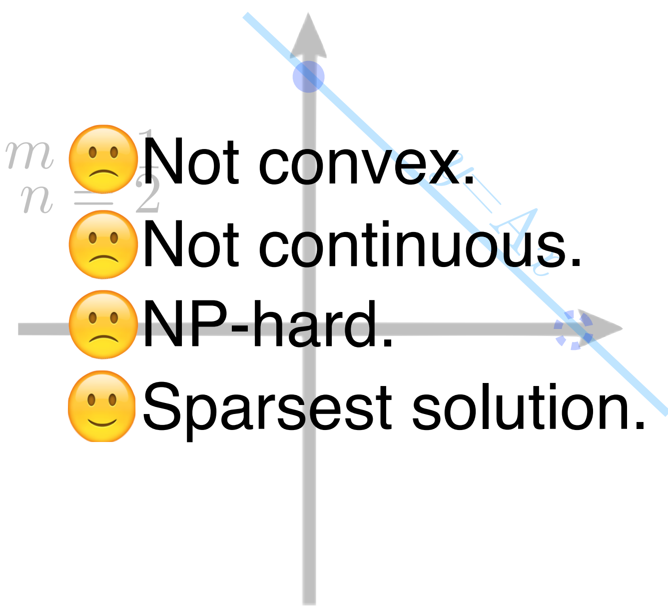
$\min_{x \in \mathbb{R}^n} \|x\|_{1.5} \quad$ s.t. $\quad y = \mathrm{A}x$



Relaxation

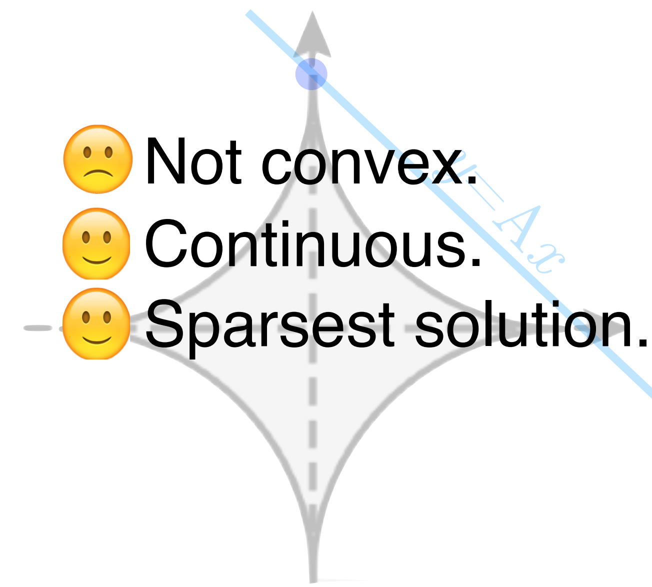
Solve $y = Ax$ where x is sparse

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



- 🙁 Not convex.
- 🙁 Not continuous.
- 🙁 NP-hard.
- 🙂 Sparsest solution.

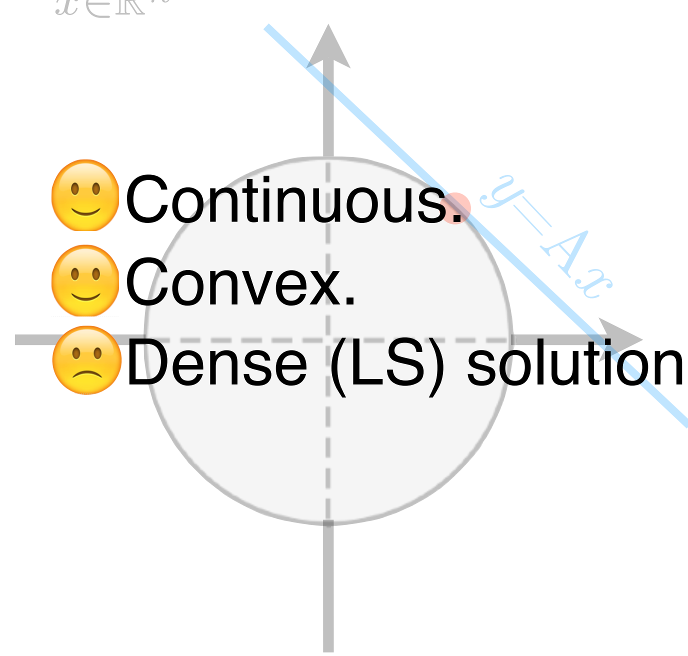
$$\min_{x\in\mathbb{R}^n} \|x\|_{0.5} \quad \text{s.t.} \quad y = Ax$$



- 🙁 Not convex.
- 🙂 Continuous.
- 🙂 Sparsest solution.

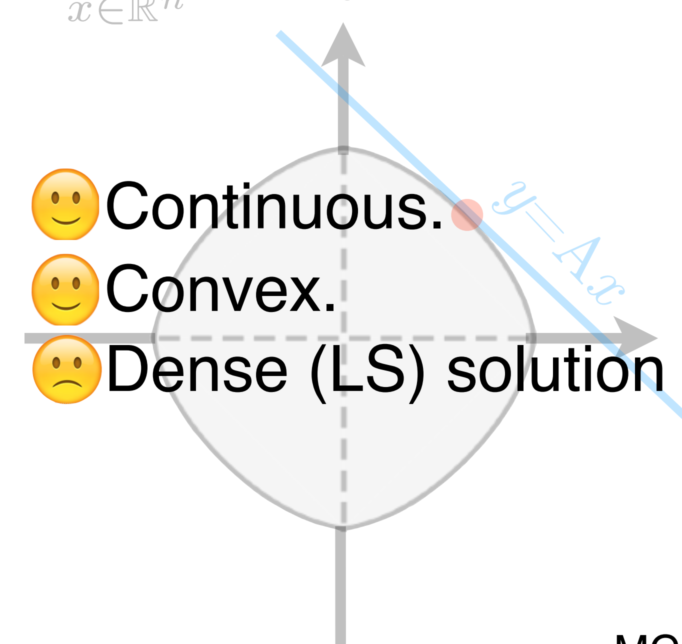
$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

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



- 🙂 Continuous.
- 🙂 Convex.
- 🙁 Dense (LS) solution

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



- 🙂 Continuous.
- 🙂 Convex.
- 🙁 Dense (LS) solution

Relaxation

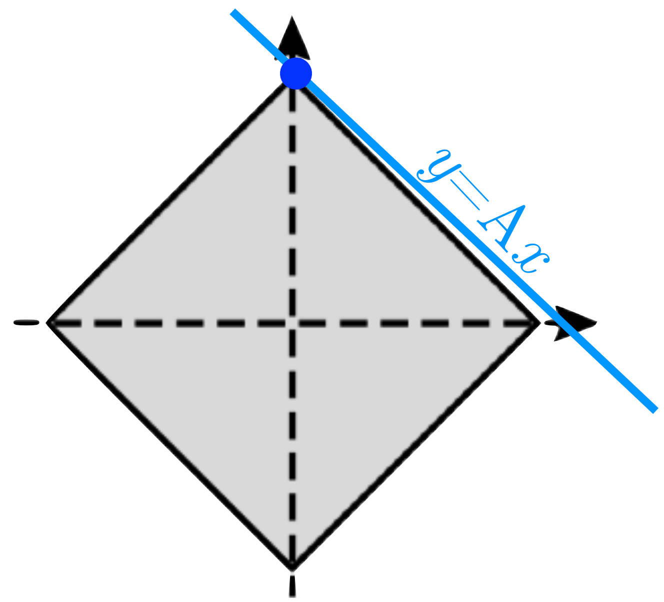
Solve $y = \mathrm{A}x$ where x is sparse

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

- Not convex.
- Not continuous.
- NP-hard.
- Sparsest solution.

Basis Pursuit

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



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

- Not convex.
- Continuous.
- Sparsest solution.

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

- Continuous.
- Convex.
- Dense (LS) solution

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

- Continuous.
- Convex.
- Dense (LS) solution

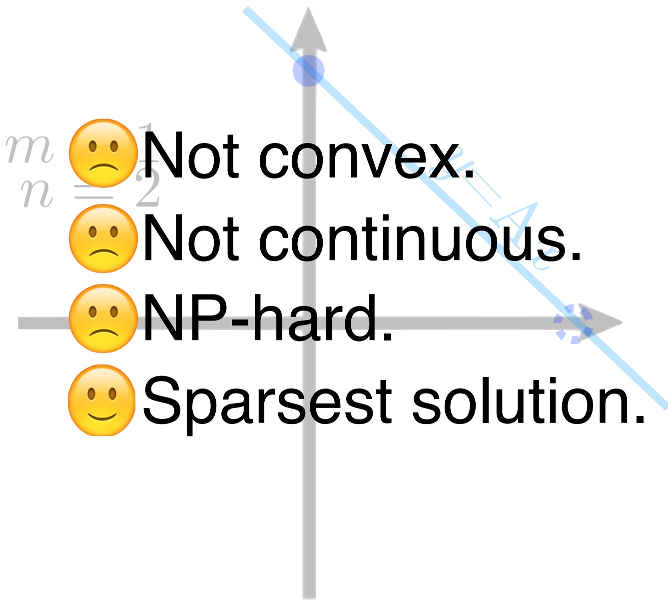
$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

Relaxation

Solve $y = Ax$ where x is sparse

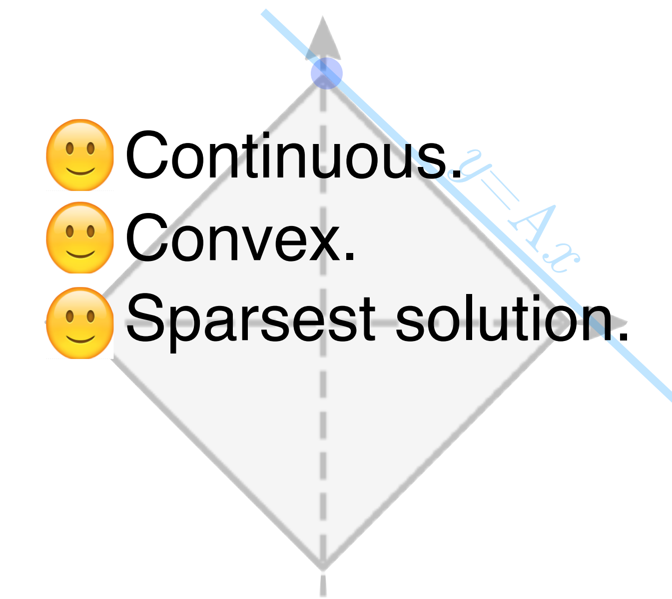
Basis Pursuit

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



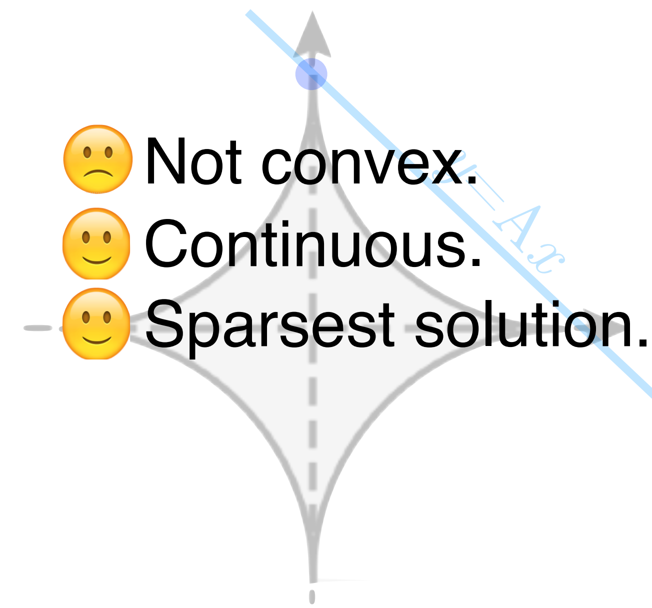
- (sad face) Not convex.
- (sad face) Not continuous.
- (sad face) NP-hard.
- (happy face) Sparsest solution.

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



- (happy face) Continuous.
- (happy face) Convex.
- (happy face) Sparsest solution.

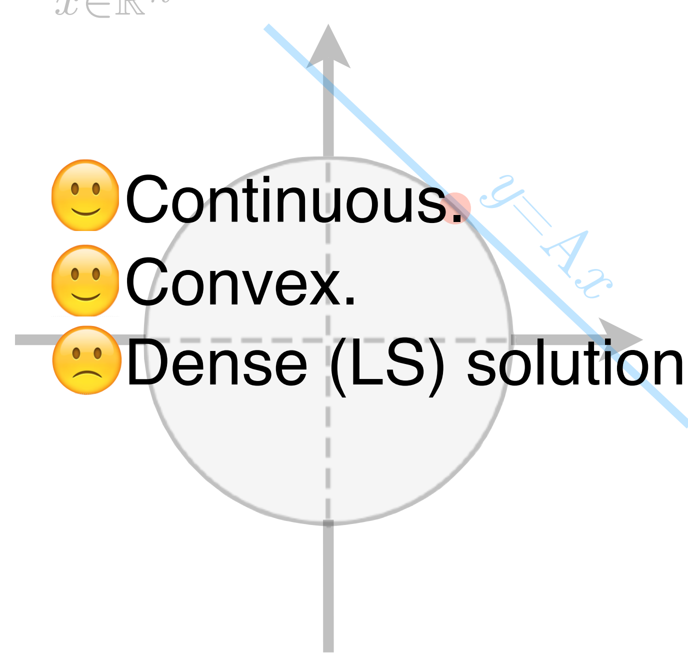
$\min_{x \in \mathbb{R}^n} \|x\|_{0.5}$ s.t. $y = Ax$



- (sad face) Not convex.
- (happy face) Continuous.
- (happy face) Sparsest solution.

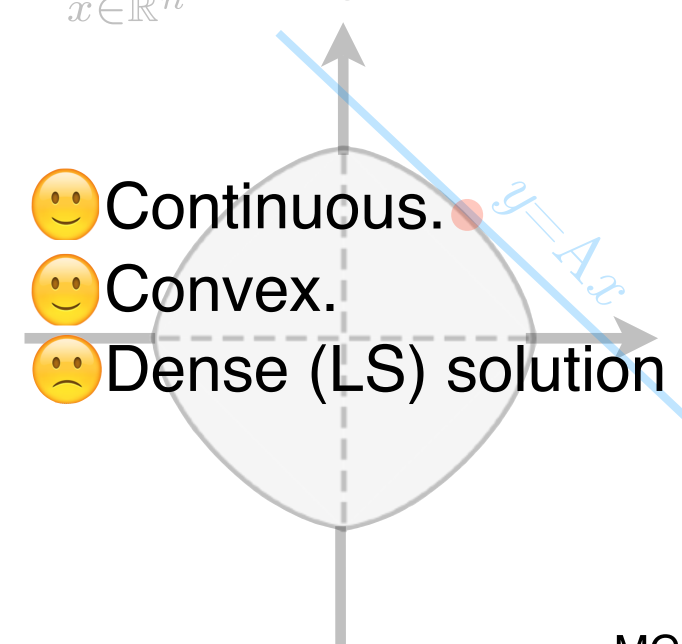
$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

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



- (happy face) Continuous.
- (happy face) Convex.
- (sad face) Dense (LS) solution

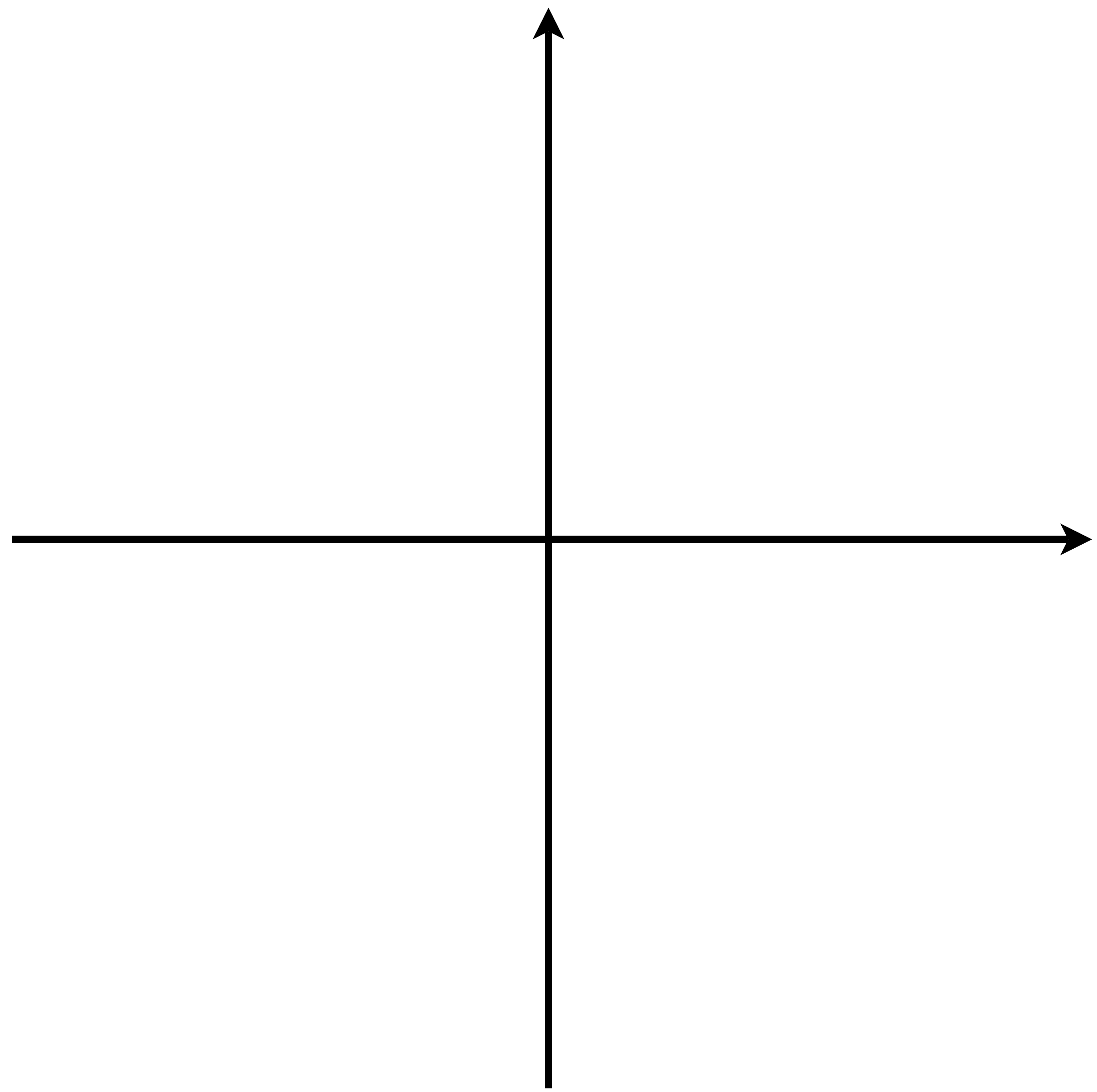
$\min_{x \in \mathbb{R}^n} \|x\|_{1.5}$ s.t. $y = Ax$



- (happy face) Continuous.
- (happy face) Convex.
- (sad face) Dense (LS) solution

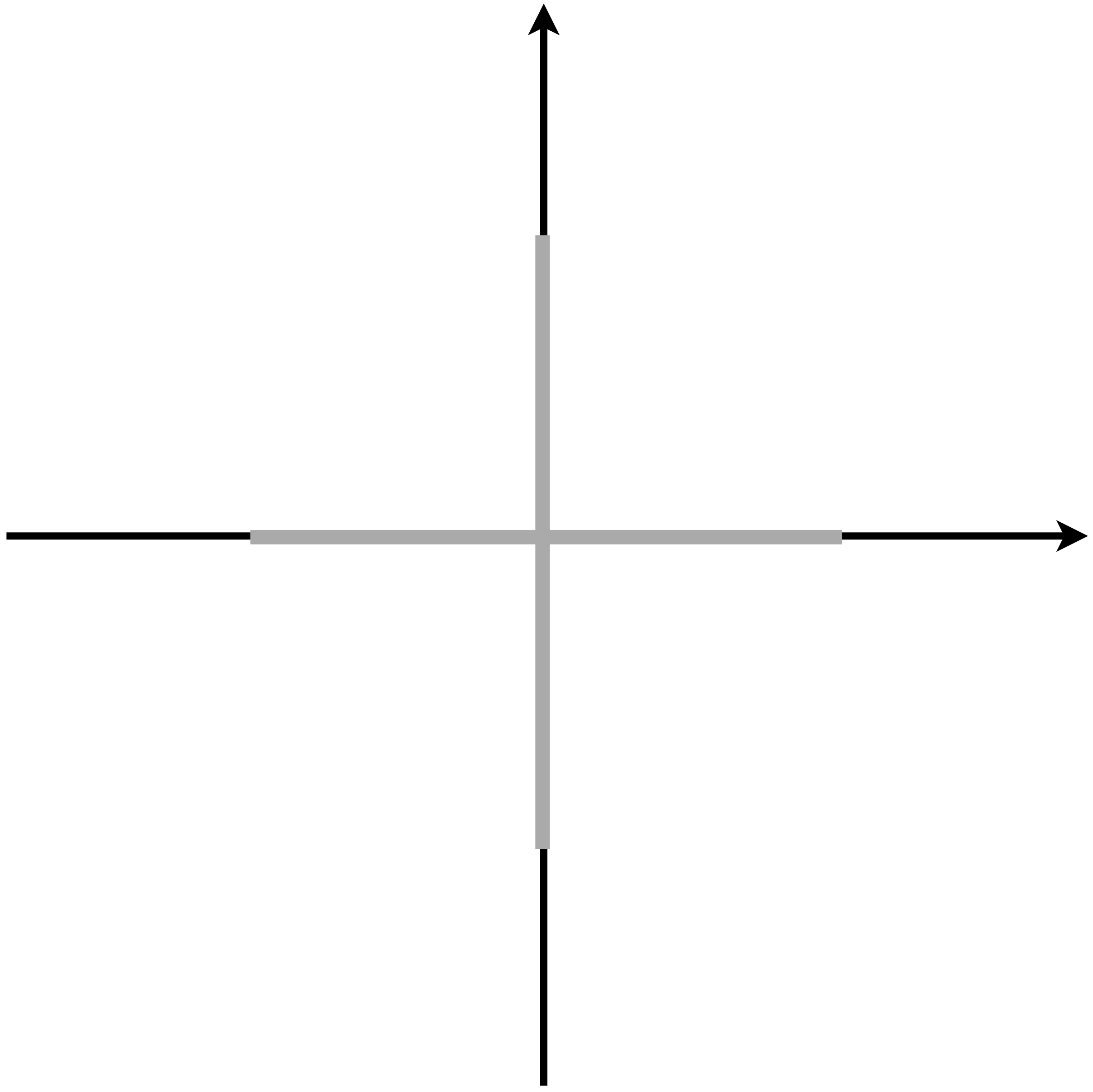
Tightest convex relaxation

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



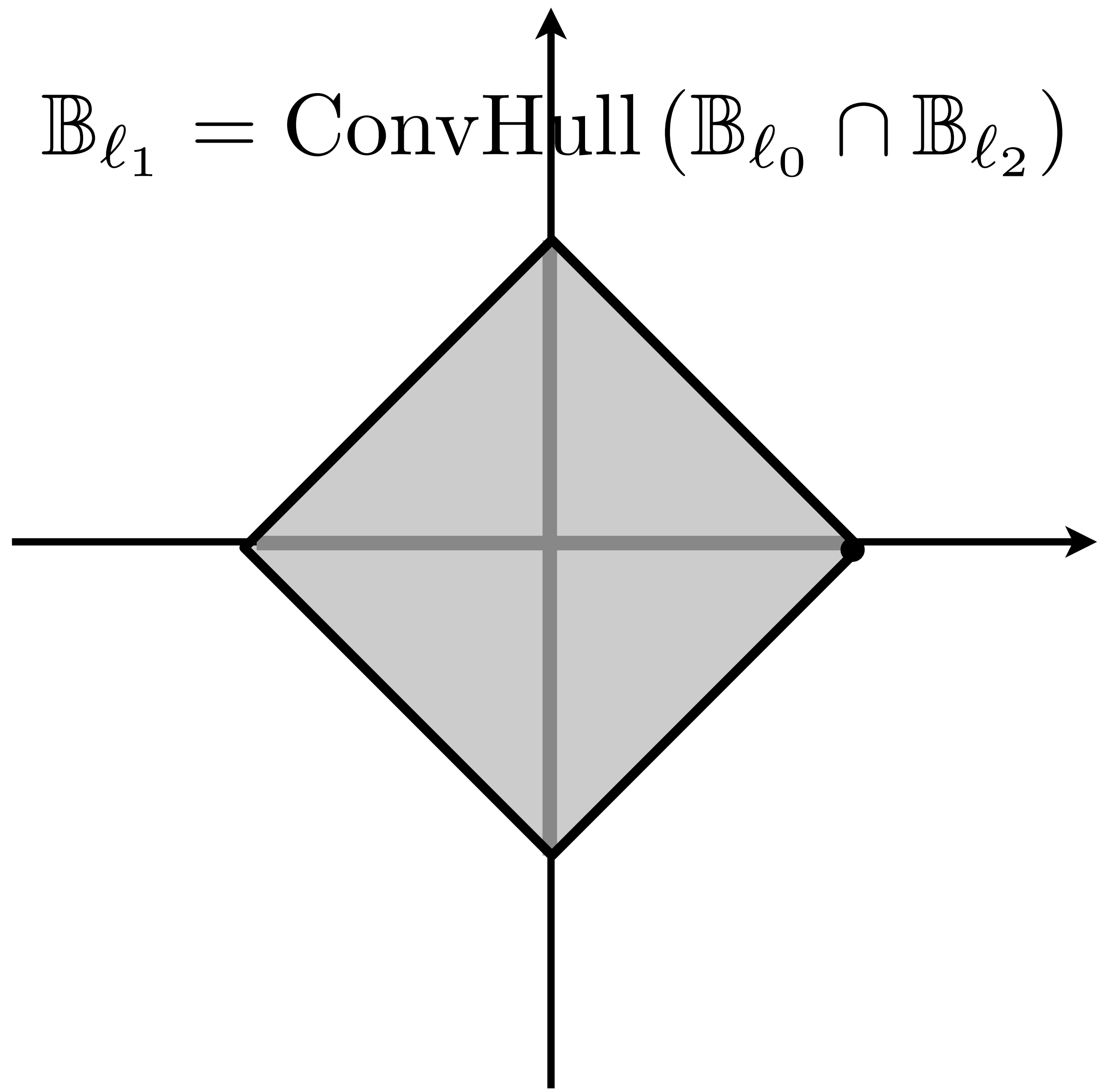
Tightest convex relaxation

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



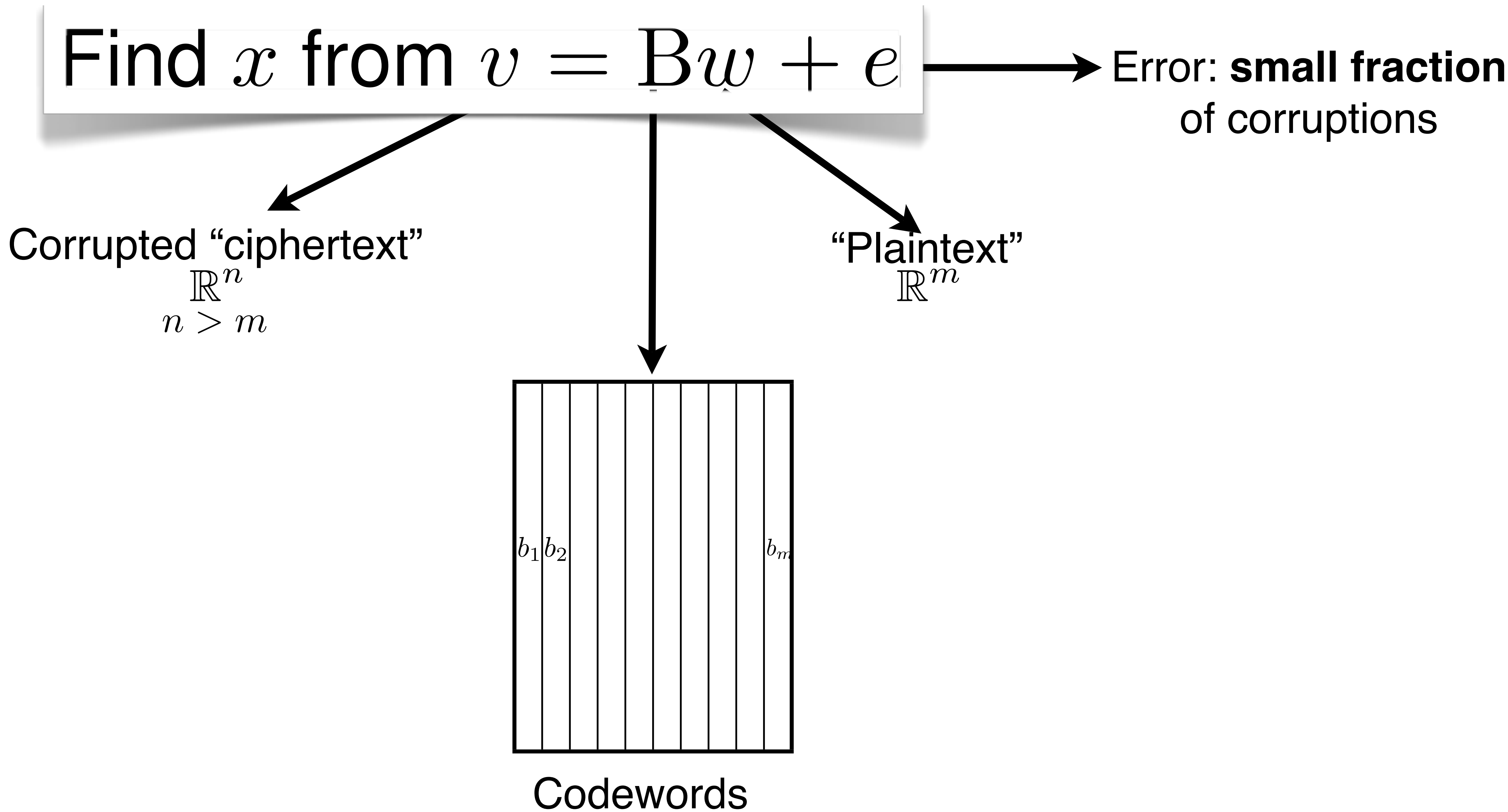
Tightest convex relaxation

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



ℓ_1 is the tightest convex relaxation of ℓ_0

Error correction problem



Error correction problem

Find x from $v = \mathrm{B}w + e$

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

$$y = \mathrm{A}v = \mathrm{A}e.$$

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

$$x^\star \in \underset{x \in \mathbb{R}^n}{\mathrm{Argmin}} \ \|x\|_1 \quad \text{s.t.} \quad y = \mathrm{A}x$$

$$w^\star = \mathrm{B}^+ x^\star$$

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

Optimization algorithms

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

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

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

  - Use your favourite LP solvers package : Cplex, Sedumi (IP), Mosek (IP), etc..
  - Recover $x^\star = (z_i^\star)_{i=1}^n - (z_i^\star)_{i=n+1}^{2n}$.
  - 🙂 High accuracy.
  - 🙁 Scaling with dimension n.
- 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

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

- Noiseless case $y = \mathrm{A}x_0$:
  - 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 constrict sensing matrices s.t. the above conditions hold ? What are the optimal scalings of the problem dimensions (n, m, s) ?
  - Necessary conditions ?
  - What if x_0 is only weakly sparse ?

Sensitivity/stability guarantees

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \left\| y - \mathrm{A}x \right\|_2^2 + \lambda \left\| x \right\|_1, \ \lambda > 0$$ (BPDN/LASSO)

- Noisy case $y = \mathrm{A}x_0 + \varepsilon$:
  - Study stability of (BPDN) solution(s) to the noise ε ?
  - ℓ_2-stability :

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

$$\left\| x^\star - x_0 \right\|_2 \leq C \left\| \varepsilon \right\|_2 .$$

Sensitivity/stability guarantees

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - \mathrm{A}x\|_2^2 + \lambda \|x\|_1 , \ \lambda > 0$$ (BPDN/LASSO)

- Noisy case $y = \mathrm{A}x_0 + \varepsilon$:
  - 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 $x^\star$ of (BPDN) obeys*

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

  - Support and sign stability (more stringent) :

Theorem (Typical statement) *Under conditions XXXX, and choice $\lambda = f(\|\varepsilon\|_2, \min_{i \in \mathrm{supp}(x)} |x_i|)$, the unique solution $x^\star$ of (BPDN) obeys*

$$\mathrm{supp}(x^\star) = \mathrm{supp}(x_0) \quad \textit{and} \quad \mathrm{sign}(x^\star) = \mathrm{sign}(x_0).$$

Sensitivity/stability guarantees

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

(BPDN/LASSO)

- Noisy case $y = \mathrm{A}x_0 + \varepsilon$:
  - 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 $x^\star$ of (BPDN) obeys*

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

  - Support and sign stability (more stringent) :

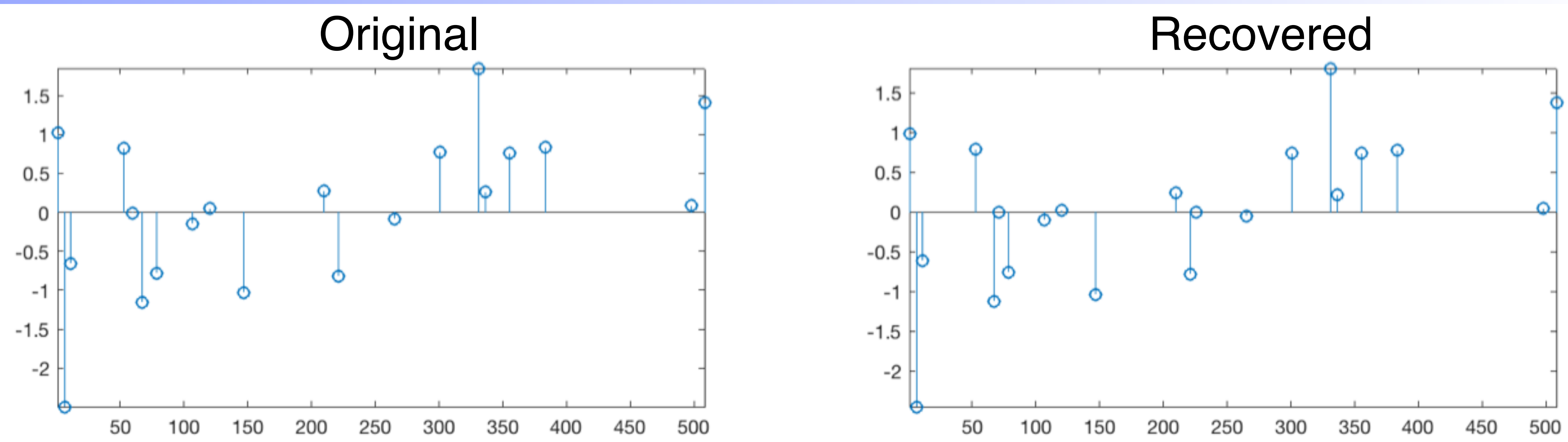
Theorem (Typical statement) *Under conditions XXXX, and choice $\lambda = f(\|\varepsilon\|_2, \min_{i \in \mathrm{supp}(x)} |x_i|)$, the unique solution $x^\star$ of (BPDN) obeys*

$$\mathrm{supp}(x^\star) = \mathrm{supp}(x_0) \quad \textit{and} \quad \mathrm{sign}(x^\star) = \mathrm{sign}(x_0).$$

- Again uniform vs non-uniform guarantees.
- Sample complexity bounds (random settings) : can we constrict sensing matrices s.t. the above conditions hold ? What are the optimal scalings of the problem dimensions (n, m, s) ?
- Necessary conditions ?
- What if x_0 is only weakly sparse ?

Sensitivity/stability guarantees

Stable support

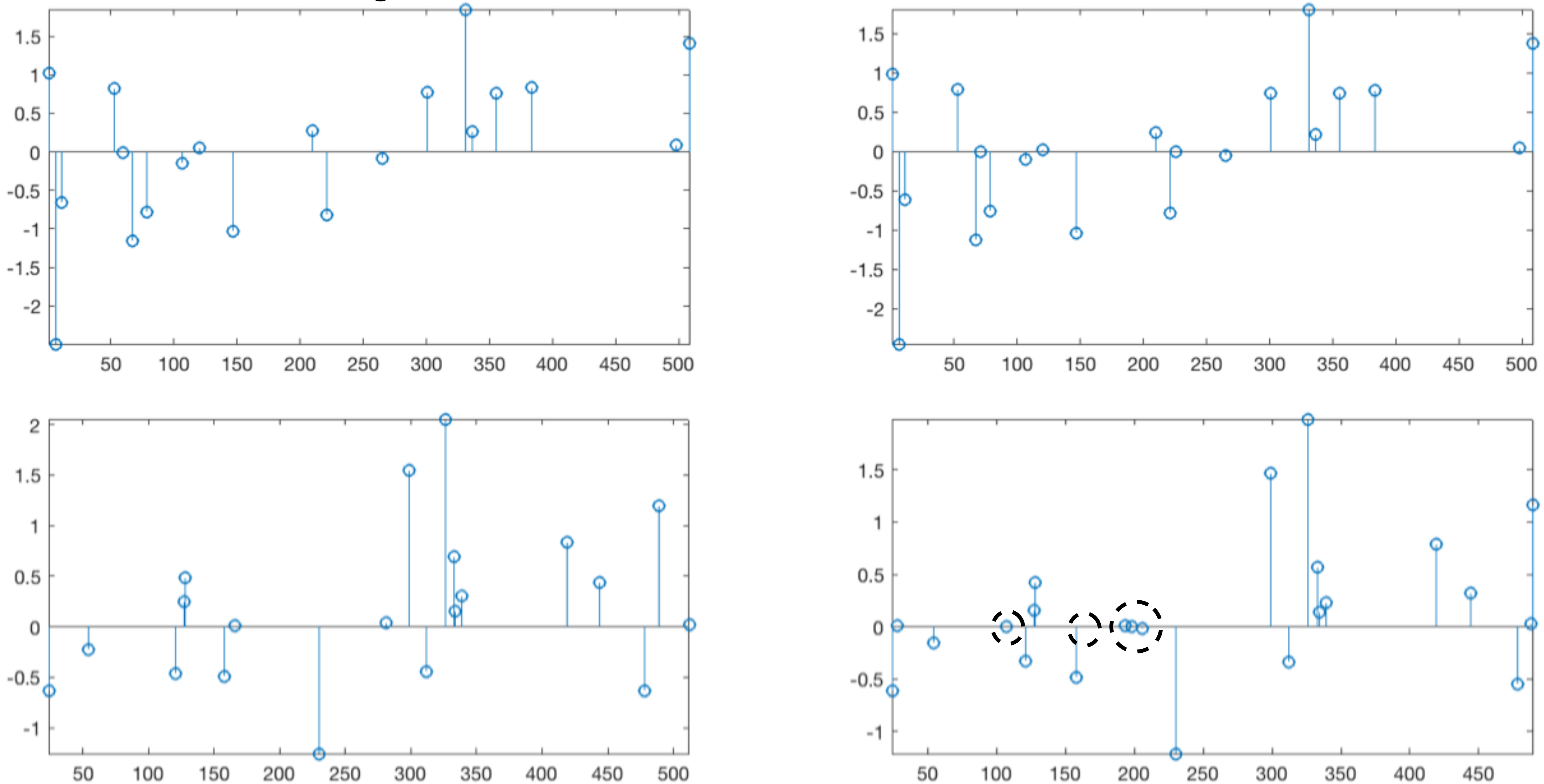


Sensitivity/stability guarantees

Original — Recovered

Stable support

No stable support but ℓ_2 stability

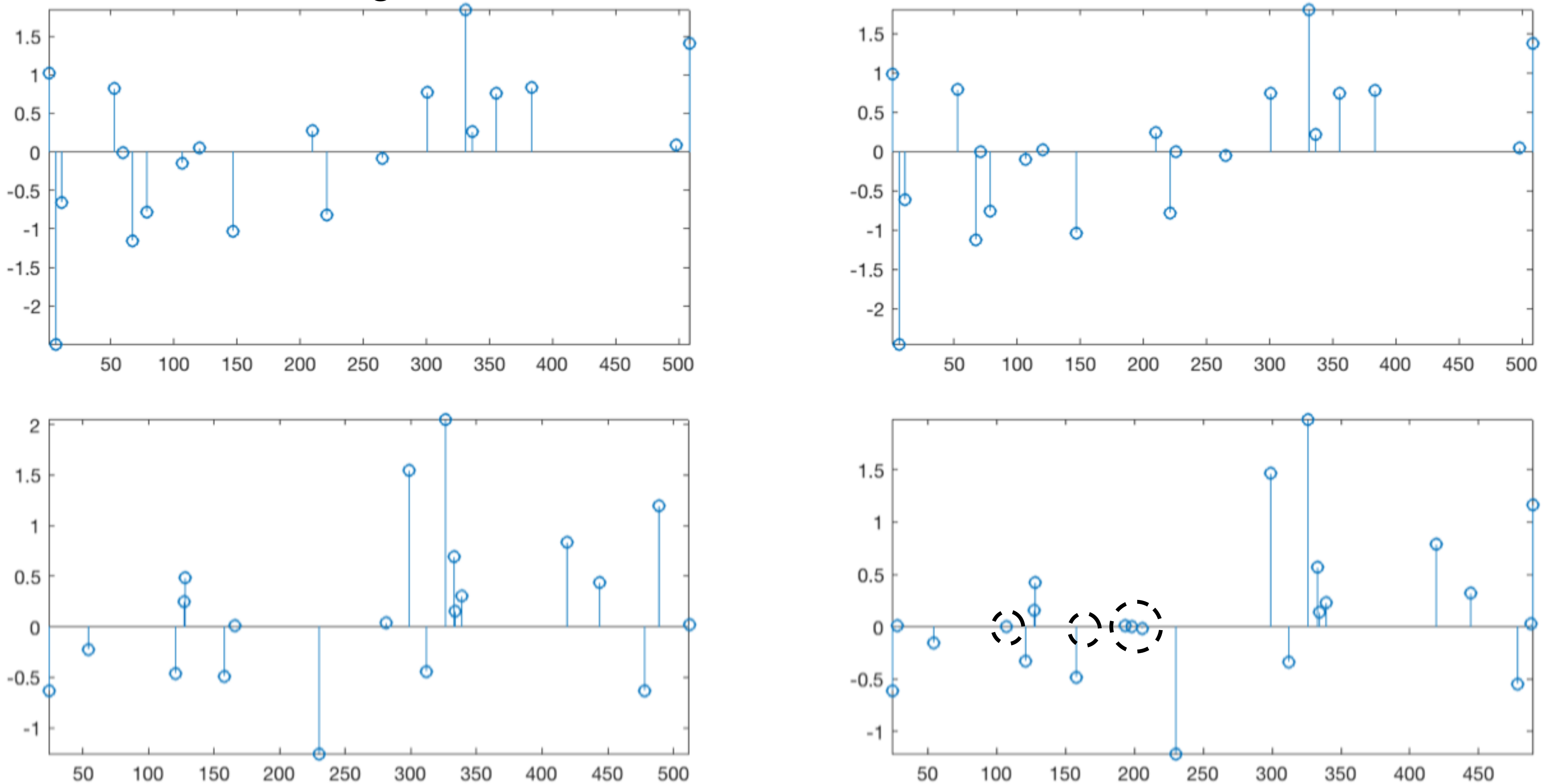


Sensitivity/stability guarantees

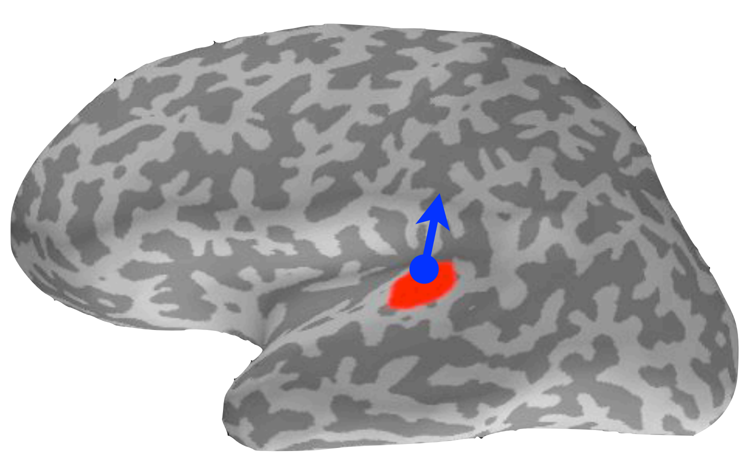
Original | Recovered

Stable support

No stable support but ℓ_2 stability



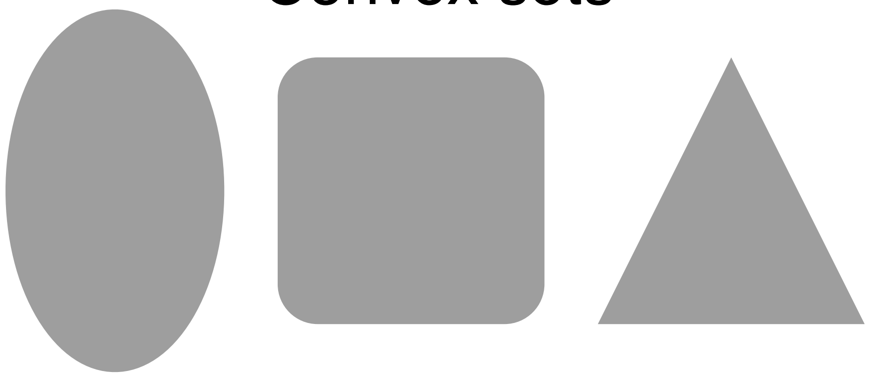
In some applications, what matters is stability of the support



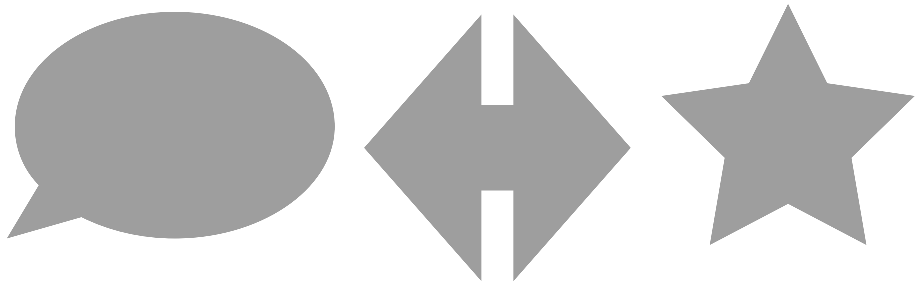
Guarantees from a geometrical perspective

Notions of convex analysis

Convex sets

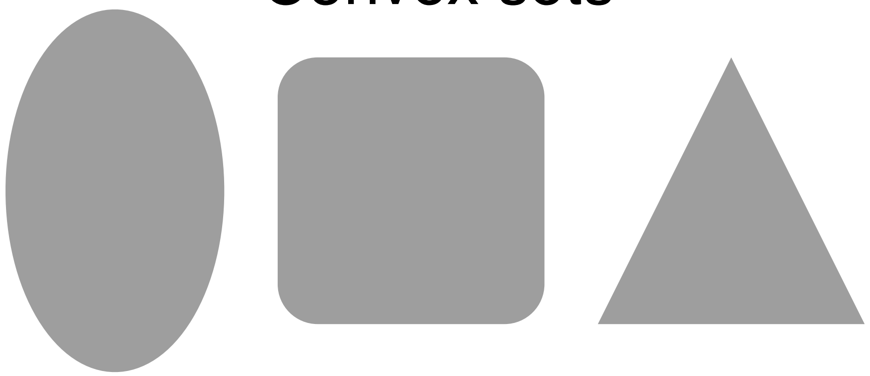


Non-convex sets

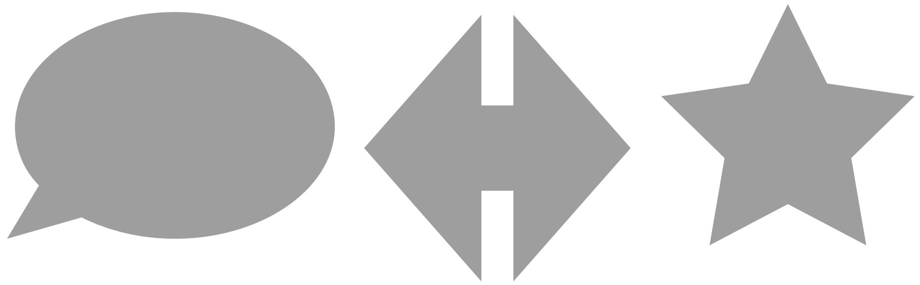


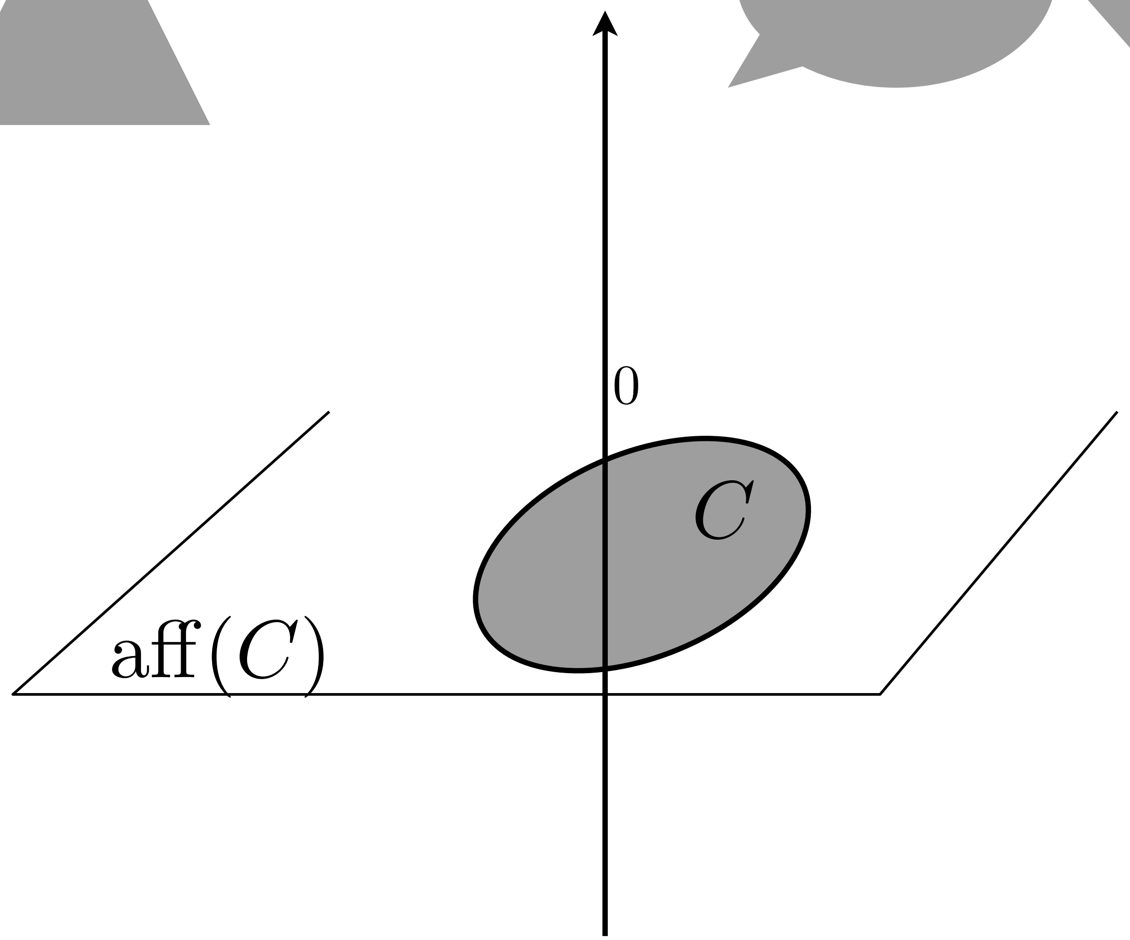
Notions of convex analysis

Convex sets



Non-convex sets

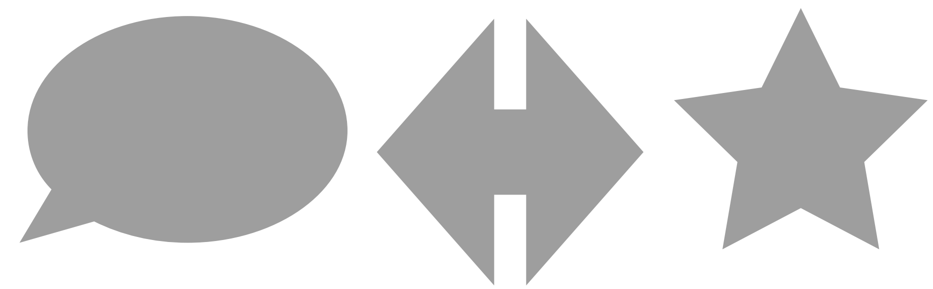


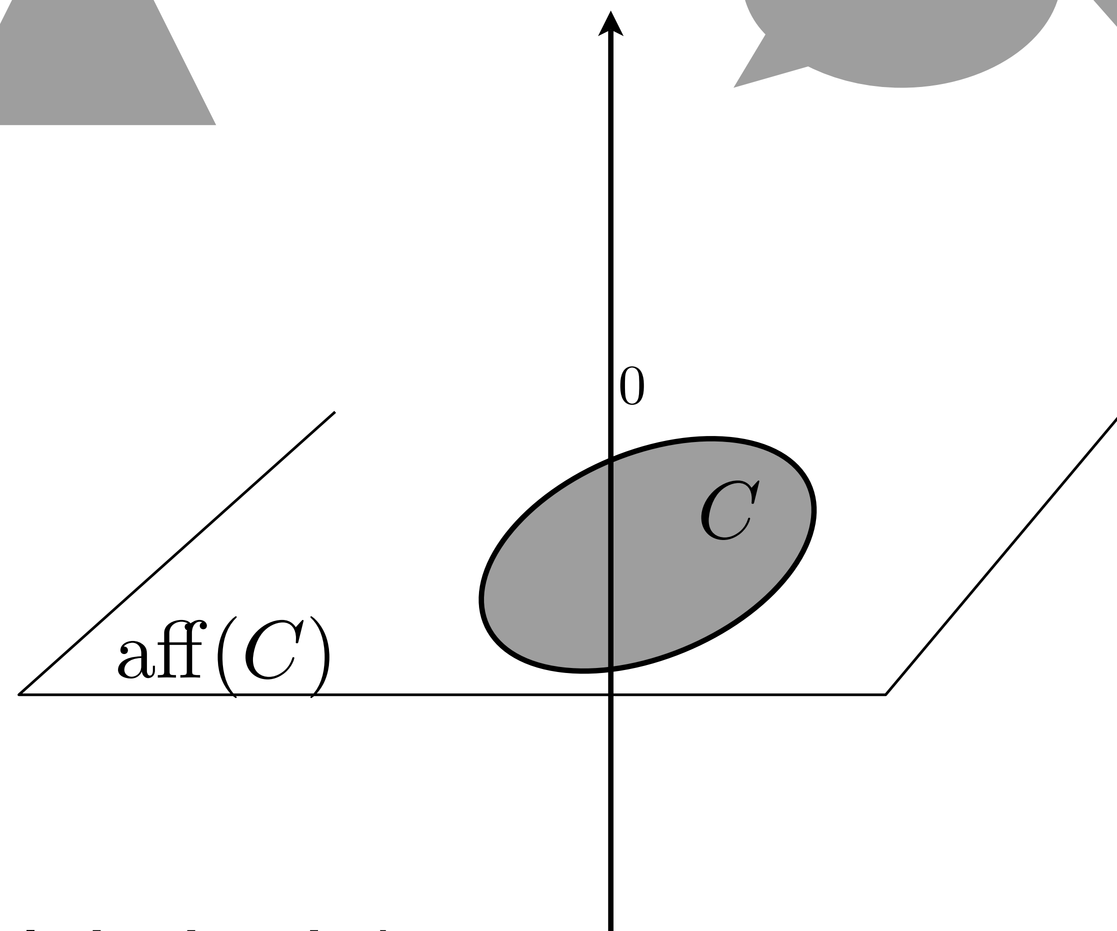


Notions of convex analysis

Convex sets

Non-convex sets





Definition (Relative interior)

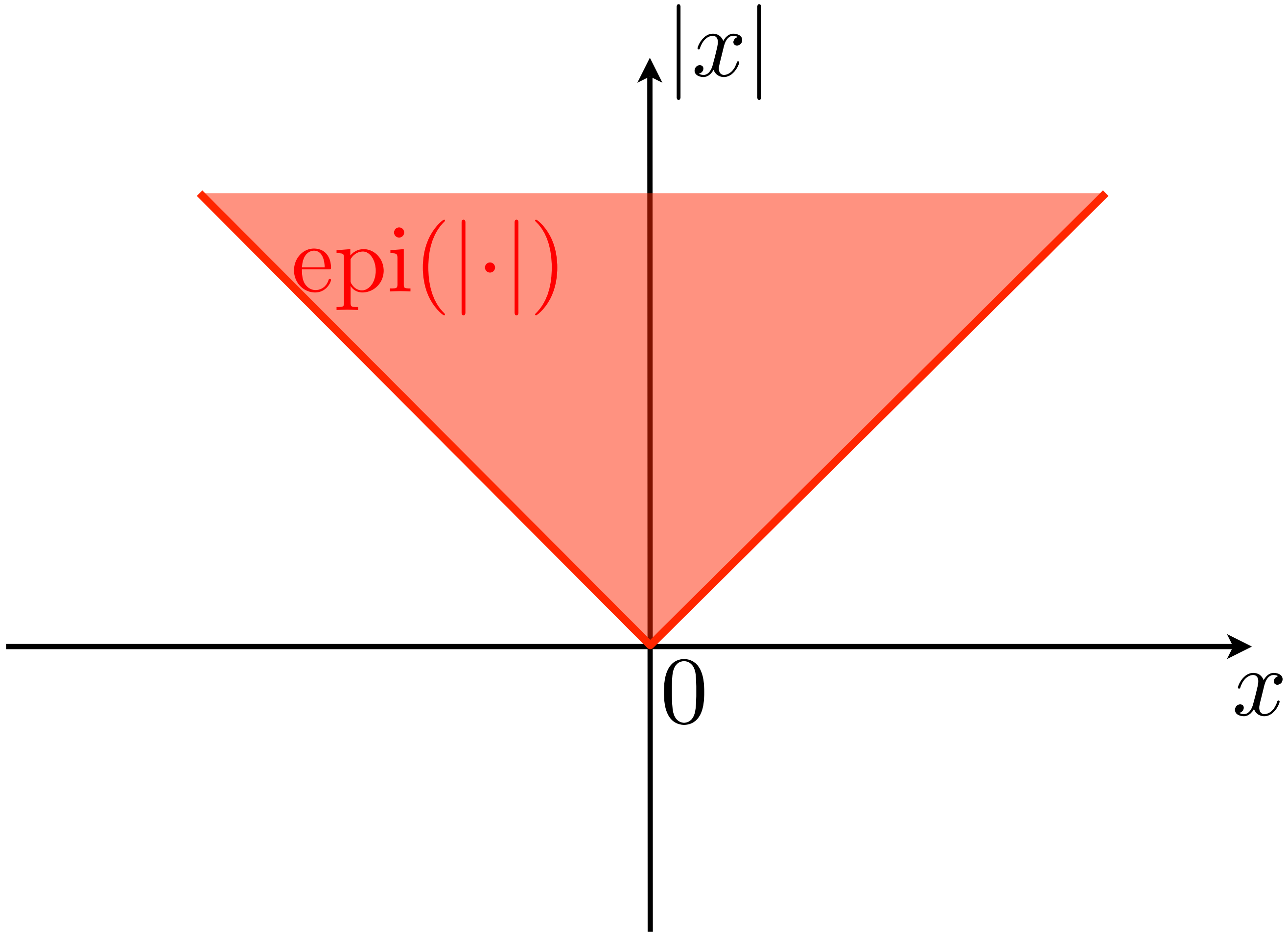
The relative interior $\mathrm{ri}(C)$ *of a convex set* C *is its interior relative to* $\mathrm{aff}(C)$*.*

C	$\mathrm{aff}(C)$	$\mathrm{ri}(C)$
$\{x\}$	$\{x\}$	$\{x\}$
$[x, x']$	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.*

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

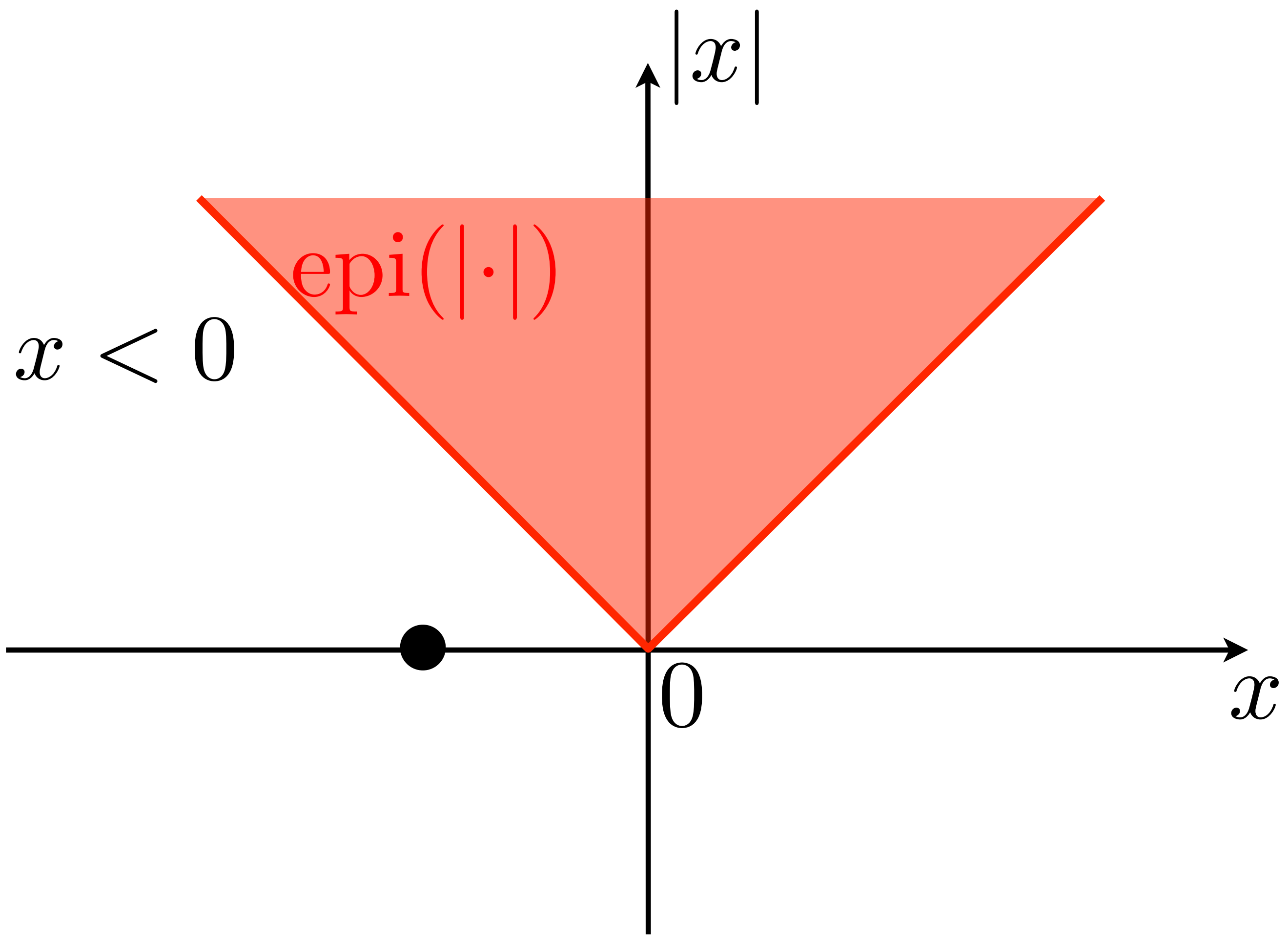


Subdifferential

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.*

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

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



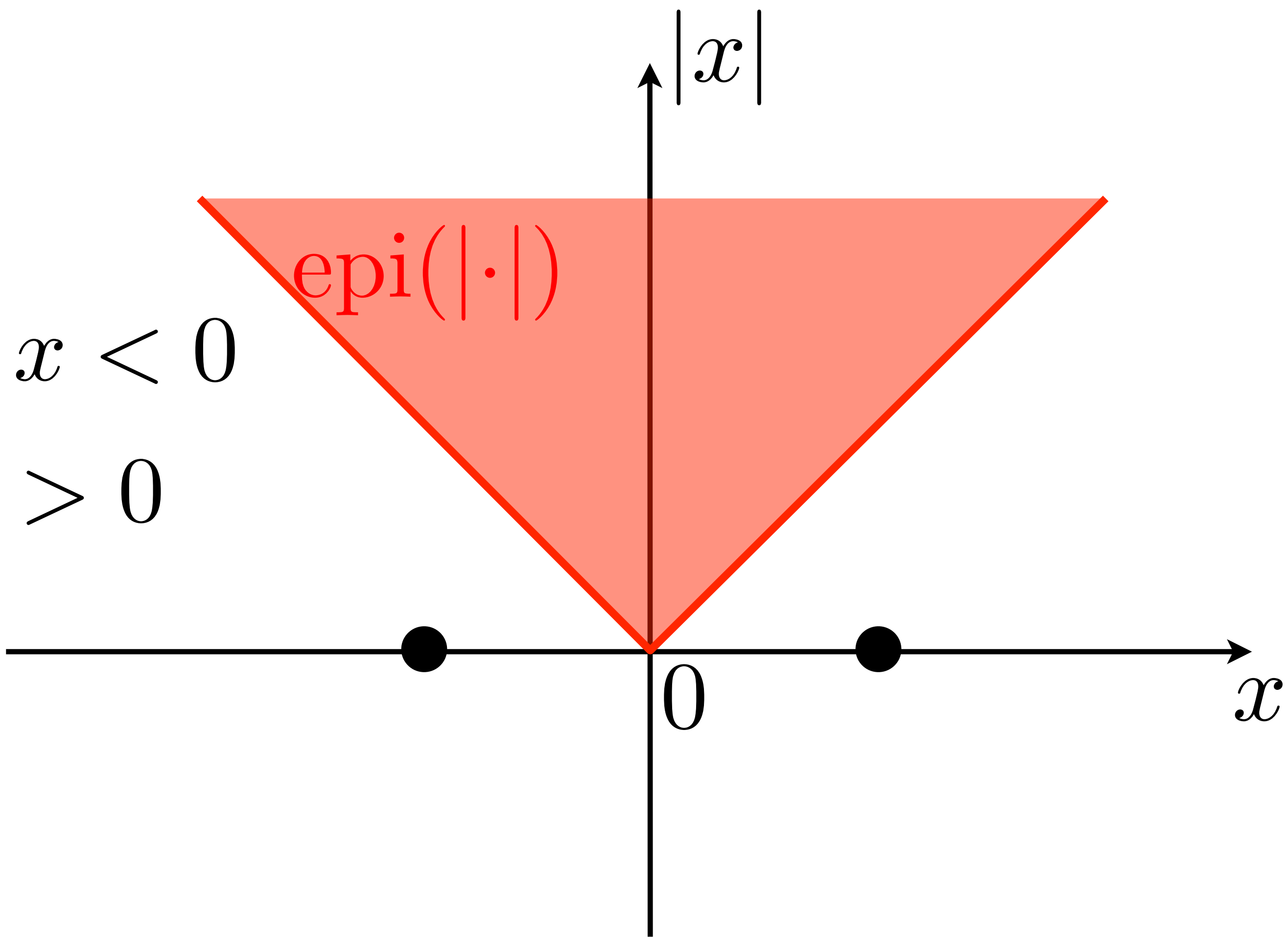
Subdifferential

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.*

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

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

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



Subdifferential

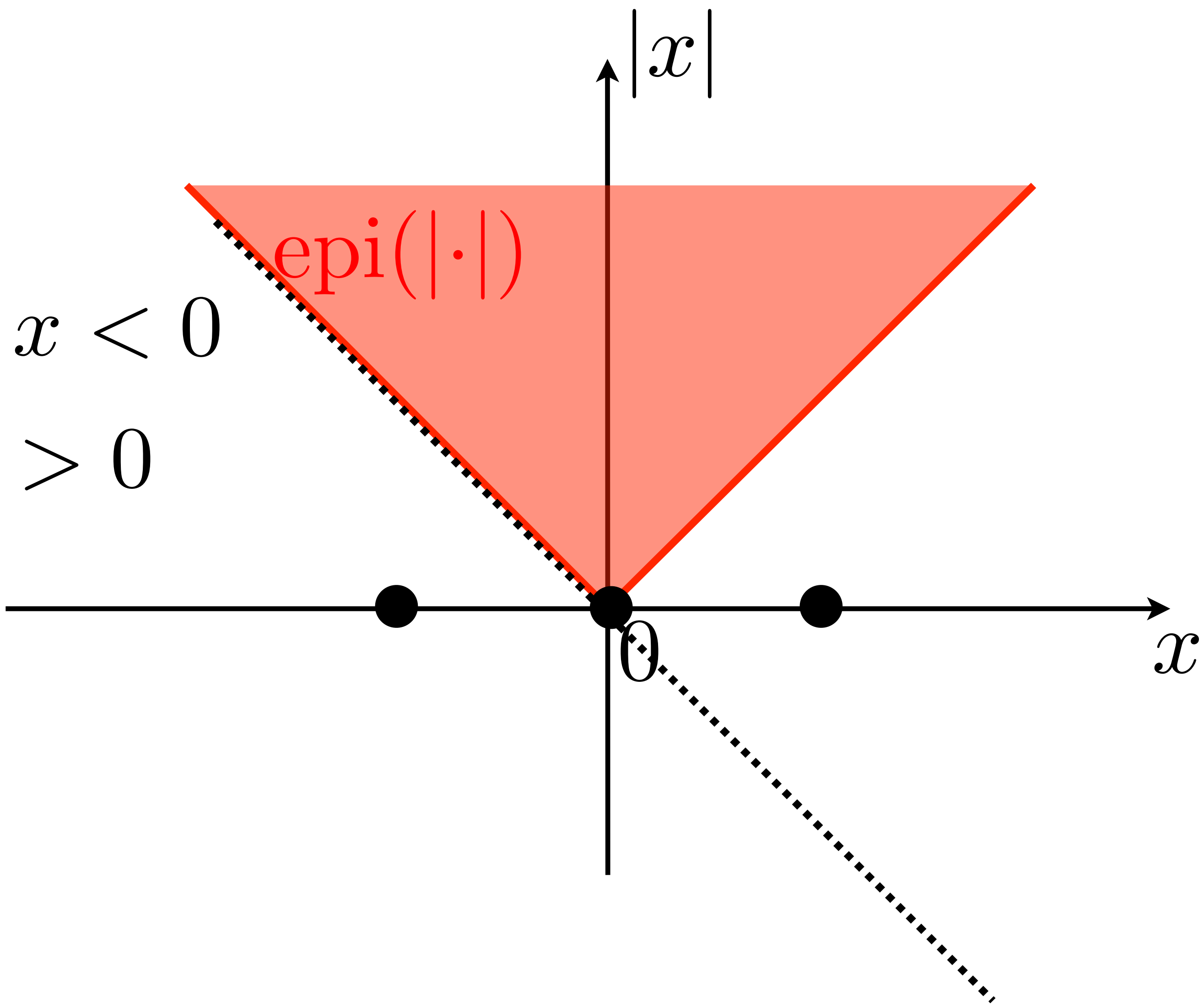
Definition (Subdifferential) *The subdifferential of a convex function at $x \in \mathbb{R}^n$ is the set of slopes of affine functions minorizing f at x, i.e.*

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

$$\partial f(x) = \{-1\}, \; x < 0$$

$$\partial f(x) = \{1\}, \; x > 0$$

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



Subdifferential

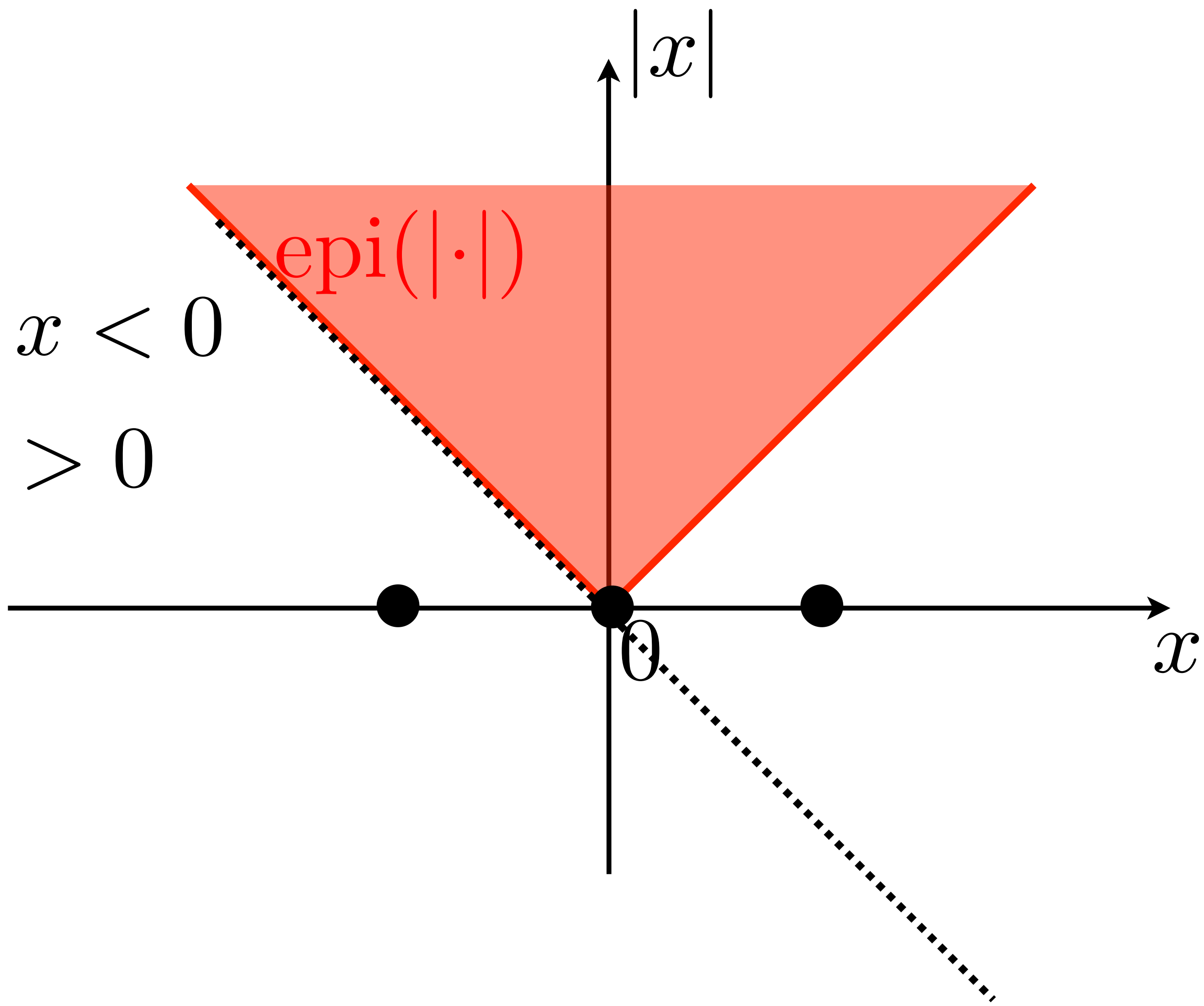
Definition (Subdifferential) *The subdifferential of a convex function at $x \in \mathbb{R}^n$ is the set of slopes of affine functions minorizing f at x, i.e.*

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

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

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

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



$$\partial \left\|x\right\|_1 = \times_{i=1}^{n} \partial |x_i| \qquad I \stackrel{\text{def}}{=} \mathrm{supp}(x)$$

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

Normal cone

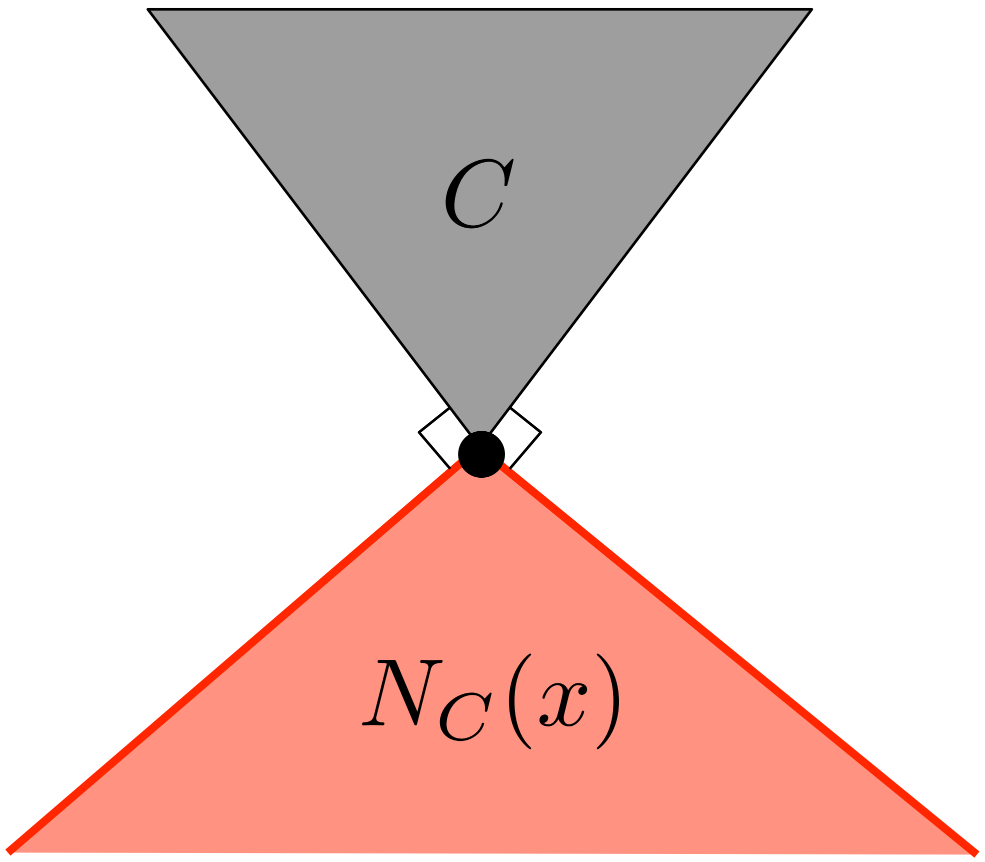
Definition (Normal cone) *The normal cone to a set C at $x \in C$ is*

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

Normal cone

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

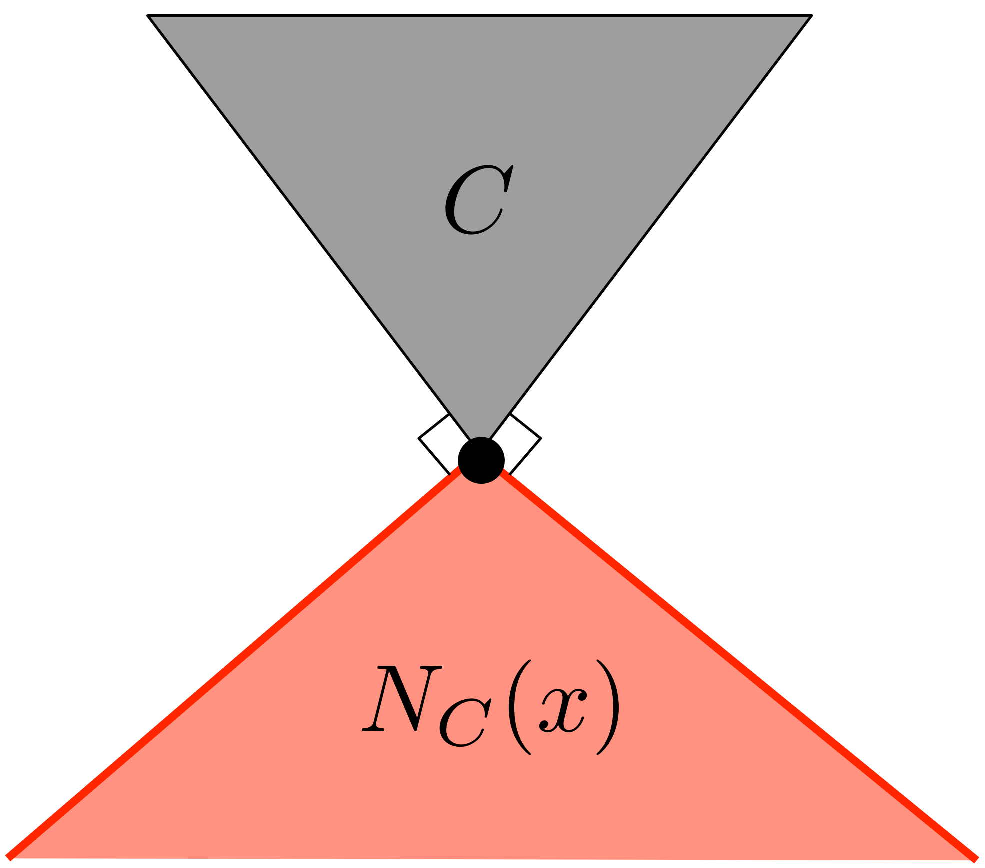
$$N_C(x) = \{u \in \mathbb{R}^n : \langle u, z - x \rangle \leq 0, \forall z \in C\}\,.$$

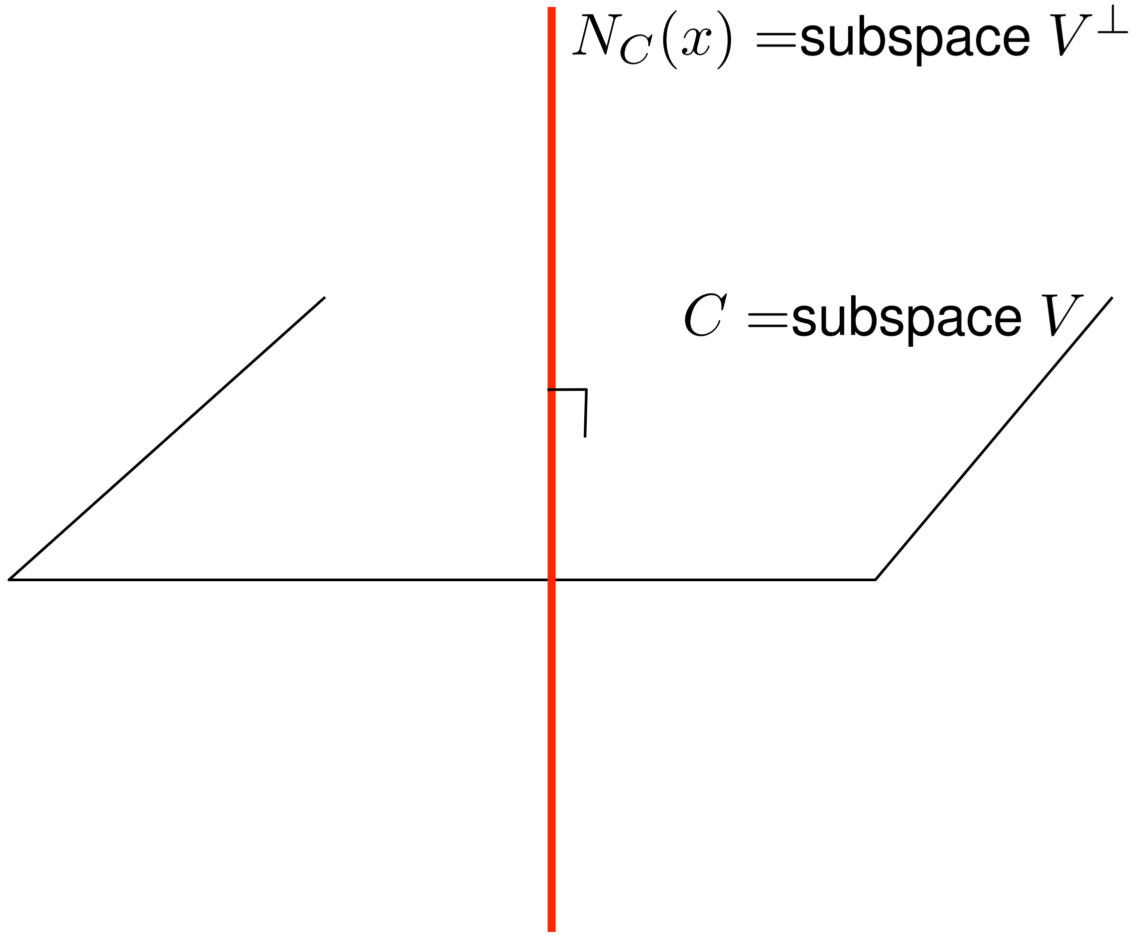


Normal cone

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

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





Optimality conditions for (BP)

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

Optimality conditions for (BP)

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

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

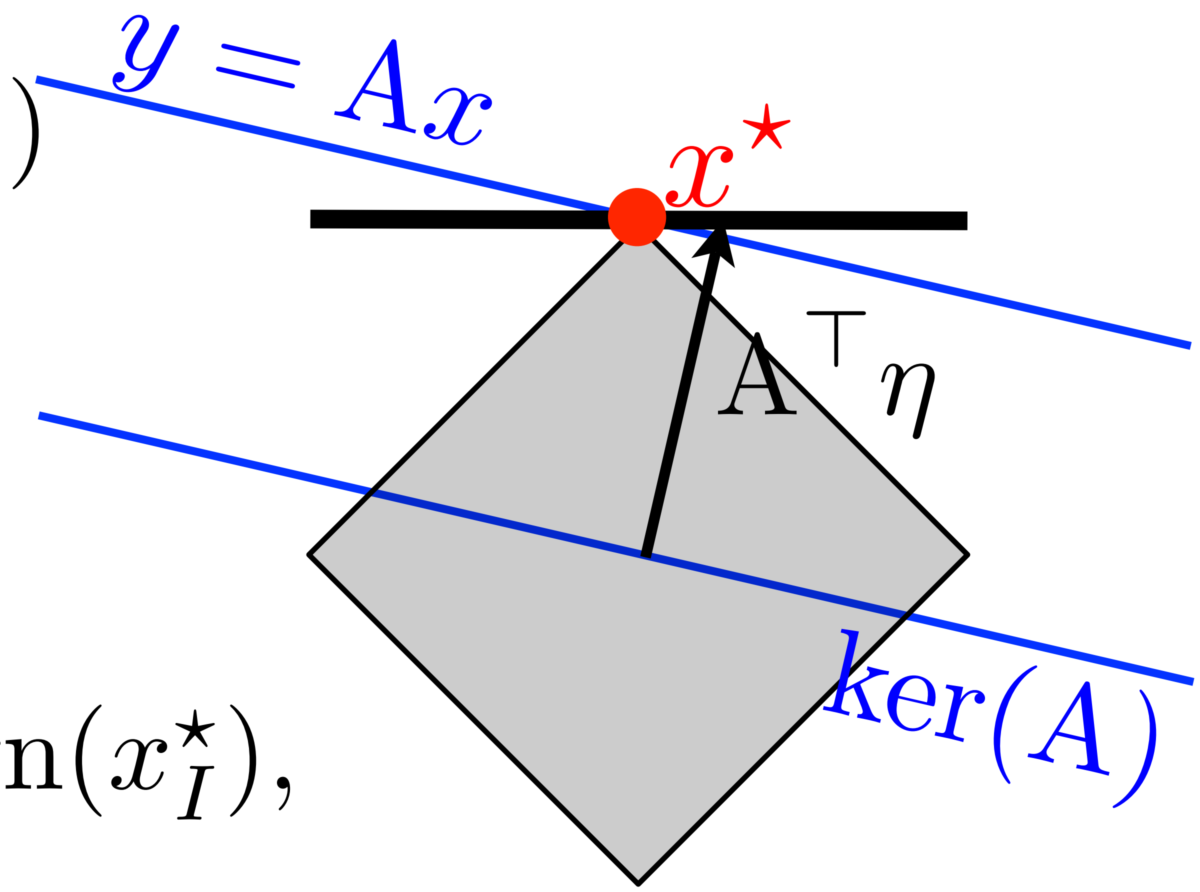
$$\iff 0 \in \partial \|x^\star\|_1 + N_{\ker(A)}(x^\star)$$

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

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

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

$$I \overset{\text{def}}{=} \text{supp}(x^\star)$$

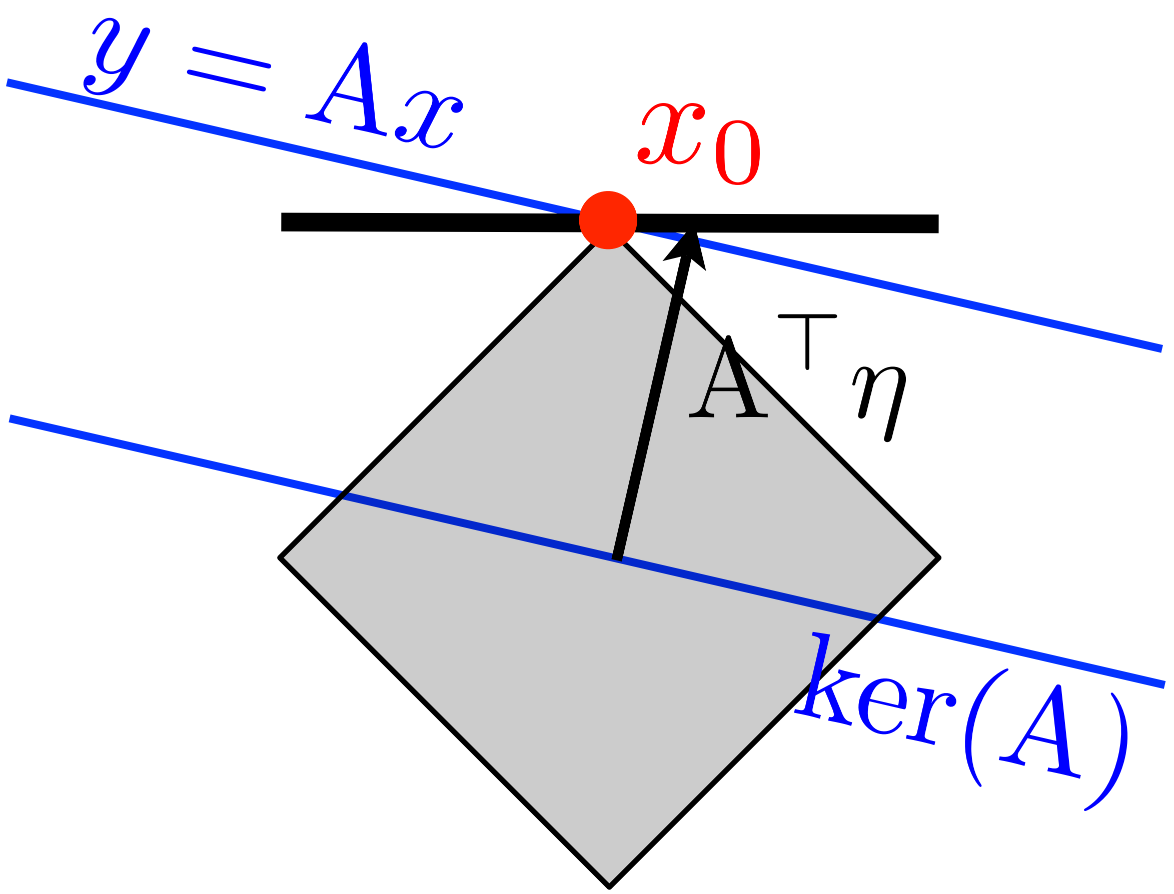


Dual certificate

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

$$\mathrm{A}^\top \eta \in \partial \|x_0\|_1$$

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**

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

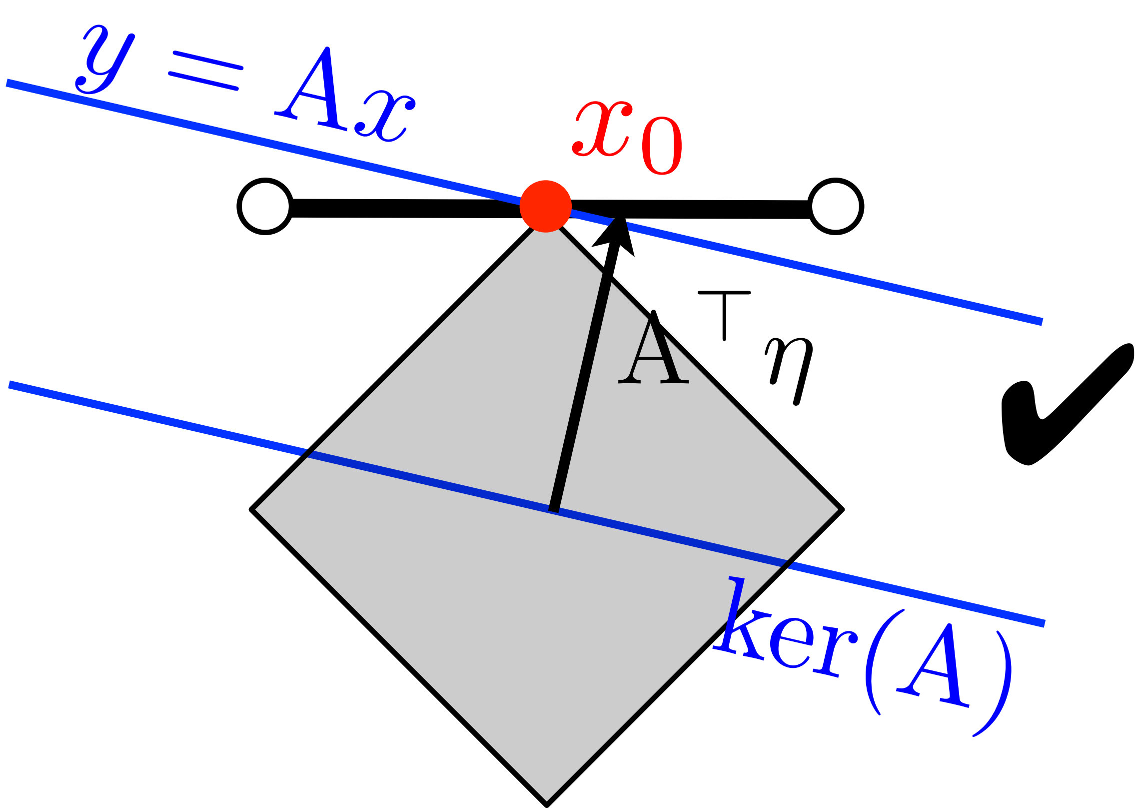
is called a **non-degenerate dual certificate**. $I \overset{\text{def}}{=} \mathrm{supp}(x_0$

Non-degenerate dual certificate

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

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

is called a **non-degenerate dual certificate**. $I \overset{\text{def}}{=} \mathrm{supp}(x_0$



Non-degenerate dual certificate

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

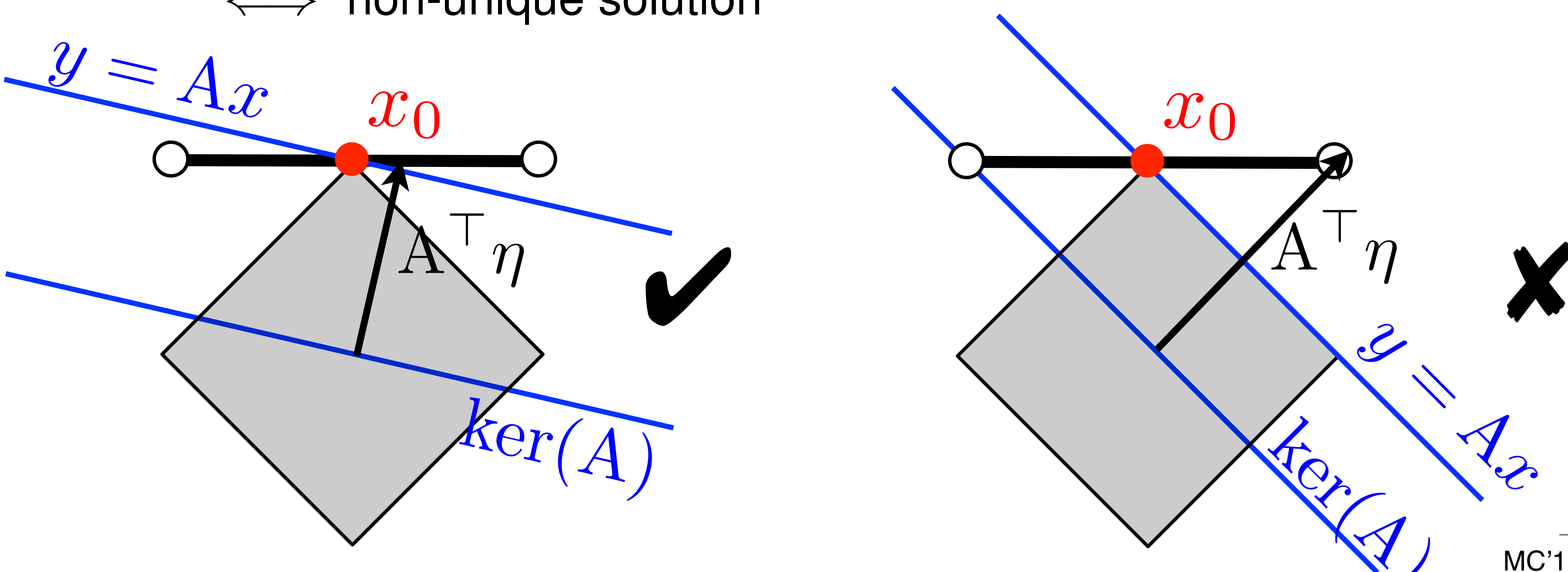
$$A^\top \eta \in \mathrm{ri}(\partial \|x_0\|_1) \iff A_I^\top \eta = \mathrm{sign}((x_0)_I) \text{ and } \|A_{I^c}^\top \eta\|_\infty < 1.$$

is called a **non-degenerate dual certificate***.* $I \overset{\text{def}}{=} \mathrm{supp}(x_0$

$A^\top \eta$ 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* $I \overset{\text{def}}{=} \mathrm{supp}(x_0)$.

- A natural assumption.
- Assume noiseless case $y = \mathrm{A}x_0$
- Assume I is known, then

$$y = \mathrm{A}x_0 = \mathrm{A}_I (x_0)_I .$$

- 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

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

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

- Even necessary when x_0 is non-trivial.

Stability without support recovery

$$y = \mathrm{A}x_0 + \varepsilon$$

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - \mathrm{A}x\|_2^2 + \lambda \|x\|_1 , \quad \lambda > 0 \qquad \text{(BPDN/LASSO)}$$

Theorem *Let $I = \mathrm{supp}(x_0)$. 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 $x^\star$ of (BPDN/LASSO) obeys*

$$\|x^\star - x_0\|_2 \leq C(c, \mathrm{A}, I, \eta) \|\varepsilon\|_2 .$$

Stability without support recovery

$$y = \mathrm{A}x_0 + \varepsilon$$

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - \mathrm{A}x\|_2^2 + \lambda \|x\|_1 \,, \;\; \lambda > 0$$ (BPDN/LASSO)

Theorem *Let $I = \mathrm{supp}(x_0)$. 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 $x^\star$ of (BPDN/LASSO) obeys*

$$\|x^\star - x_0\|_2 \leq C(c, \mathrm{A}, I, \eta)\,\|\varepsilon\|_2 \,.$$

- Even necessary when x_0 is non-trivial.

Stable support and sign recovery

$$y = \mathrm{A}x_0 + \varepsilon$$

$$\boxed{\min_{x \in \mathbb{R}^n} \tfrac{1}{2} \|y - \mathrm{A}x\|_2^2 + \lambda \|x\|_1 \,, \;\; \lambda > 0} \qquad \text{(BPDN/LASSO)}$$

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

$$\eta_F = \mathrm{A}_I (\mathrm{A}_I^\top \mathrm{A}_I)^{-1} \mathrm{sign}((x_0)_I)$$

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 $x^\star$ which moreover satisfies

$$\mathrm{supp}(x^\star) = I \;\; \textit{and} \;\; \mathrm{sign}(x^\star) = \mathrm{sign}(x_0).$$

Stable support and sign recovery

$$y = \mathrm{A}x_0 + \varepsilon$$

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|y - \mathrm{A}x\|_2^2 + \lambda \|x\|_1 , \quad \lambda > 0 \qquad \text{(BPDN/LASSO)}$$

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

$$\eta_F = \mathrm{A}_I(\mathrm{A}_I^\top \mathrm{A}_I)^{-1}\mathrm{sign}((x_0)_I)$$

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 $x^\star$ which moreover satisfies

$$\mathrm{supp}(x^\star) = I \ \textit{ and } \ \mathrm{sign}(x^\star) = \mathrm{sign}(x_0).$$

- 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 (next lecture).

https://fadili.users.greyc.fr/

Thanks
Any questions ?