ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Winter Conference in Statistics 2013

Compressed Sensing

LECTURE #1-2

Motivation & geometric insights

lions@epfl

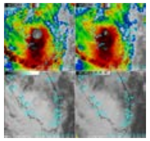
Prof. Dr. Volkan Cevher

volkan.cevher@epfl.ch

LIONS/Laboratory for Information and Inference Systems

Major trends

160MP



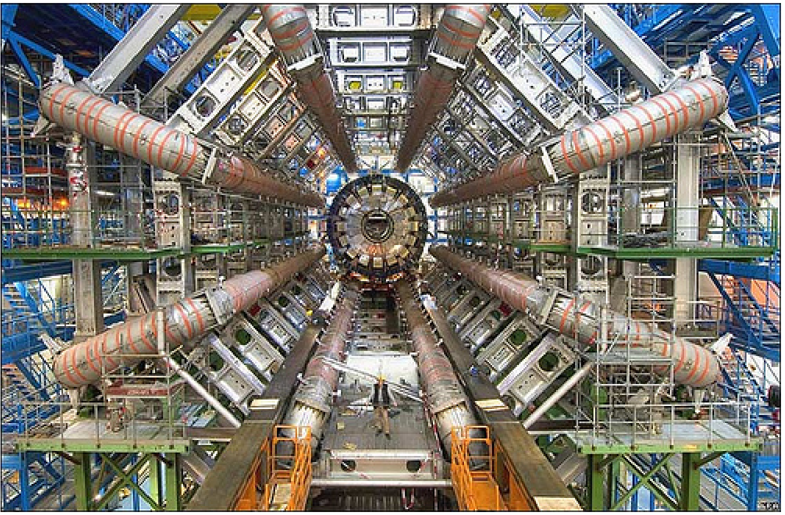
higher resolution / denser sampling

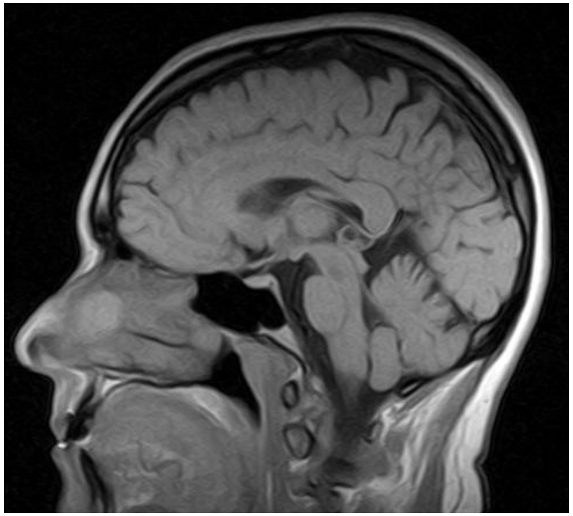
x

large numbers of sensors

x

increasing # of modalities / mobility





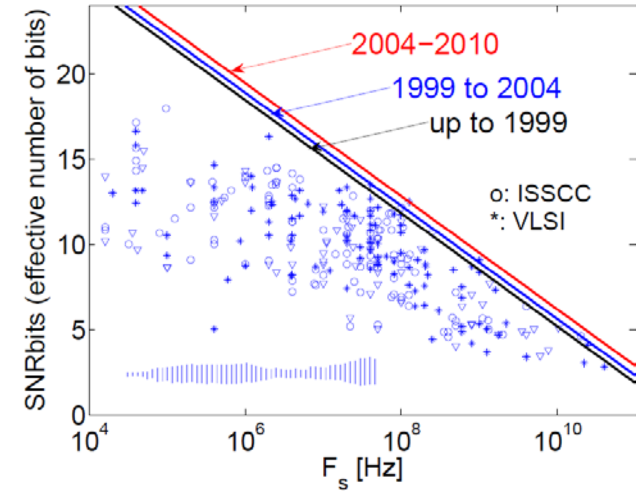
Motivation: solve bigger / more important problems

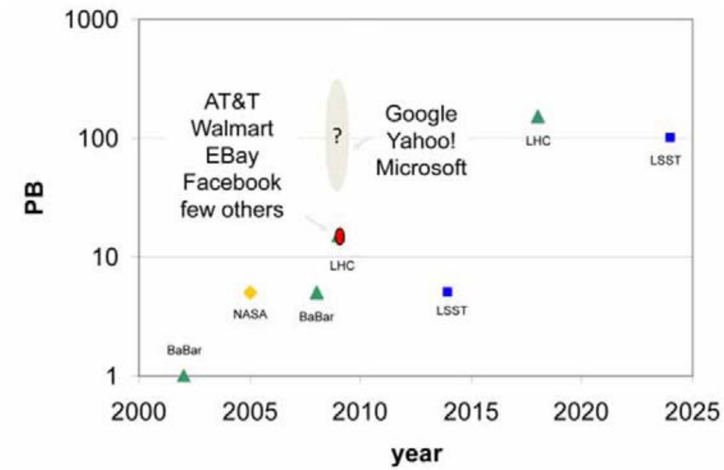
decrease acquisition times / costs

entertainment / new consumer products...

Problems of the current paradigm

- **Sampling at Nyquist rate**
  - expensive / difficult
- **Data deluge**
  - communications / storage
- **Sample then compress**
  - inefficient / impossible / not future proof

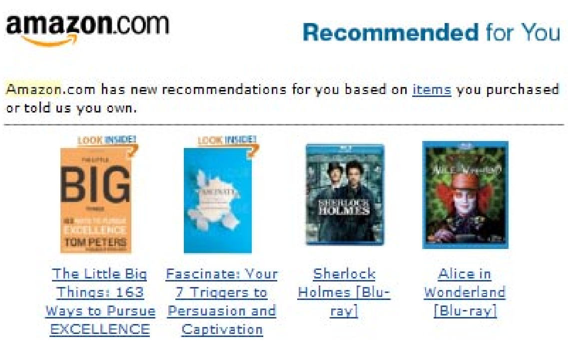




Recommended for you: A more familiar example

- Recommender systems
  - observe partial information

“ratings”
“clicks”
“purchases”
“compatibilities”



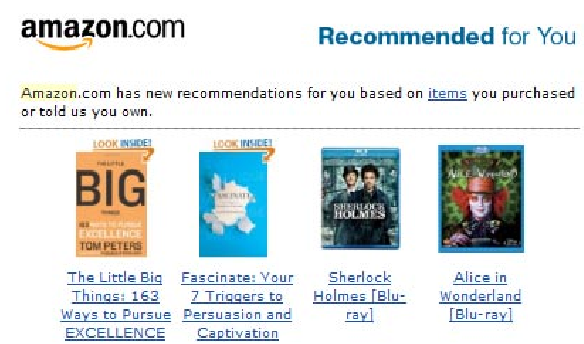
Recommended for you: A more familiar example

- Recommender systems
  - observe partial information

“ratings”
“clicks”
“purchases”
“compatibilities”

- The Netflix problem

  - from approx. 100,000,000 ratings predict 3,000,000 ratings
  - 17770 movies x 480189 users
  - how would you automatically predict?



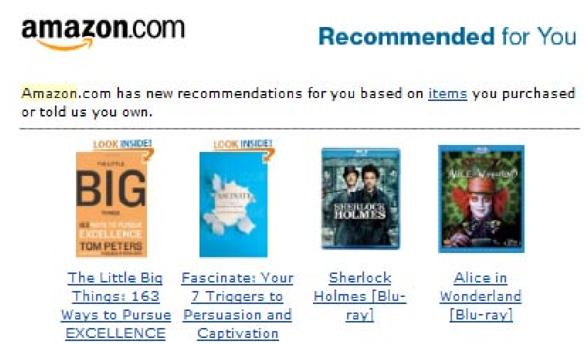
Recommended for you: A more familiar example

- Recommender systems
  - observe partial information

“ratings”
“clicks”
“purchases”
“compatibilities”

- The Netflix problem 

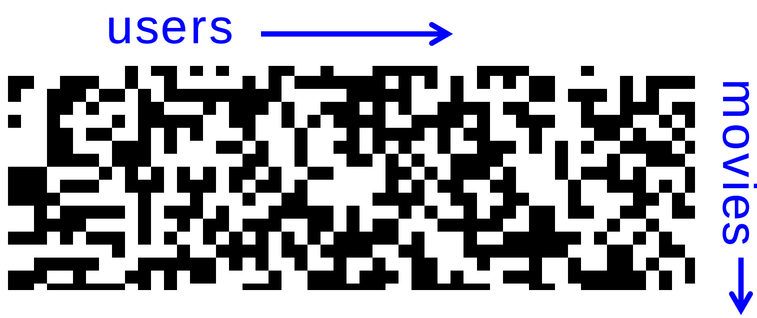
  - from approx. 100,000,000 ratings predict 3,000,000 ratings
  - 17770 movies x 480189 users
  - how would you automatically predict?
  - **what is it worth?**



Theoretical set-up

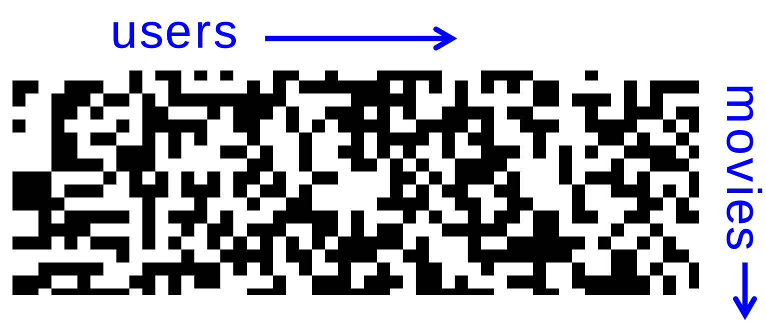
- Matrix completion for Netflix

$X =$



Theoretical set-up

- Matrix completion for Netflix

$X =$ 

- Mathematical underpinnings: ***compressive sensing***

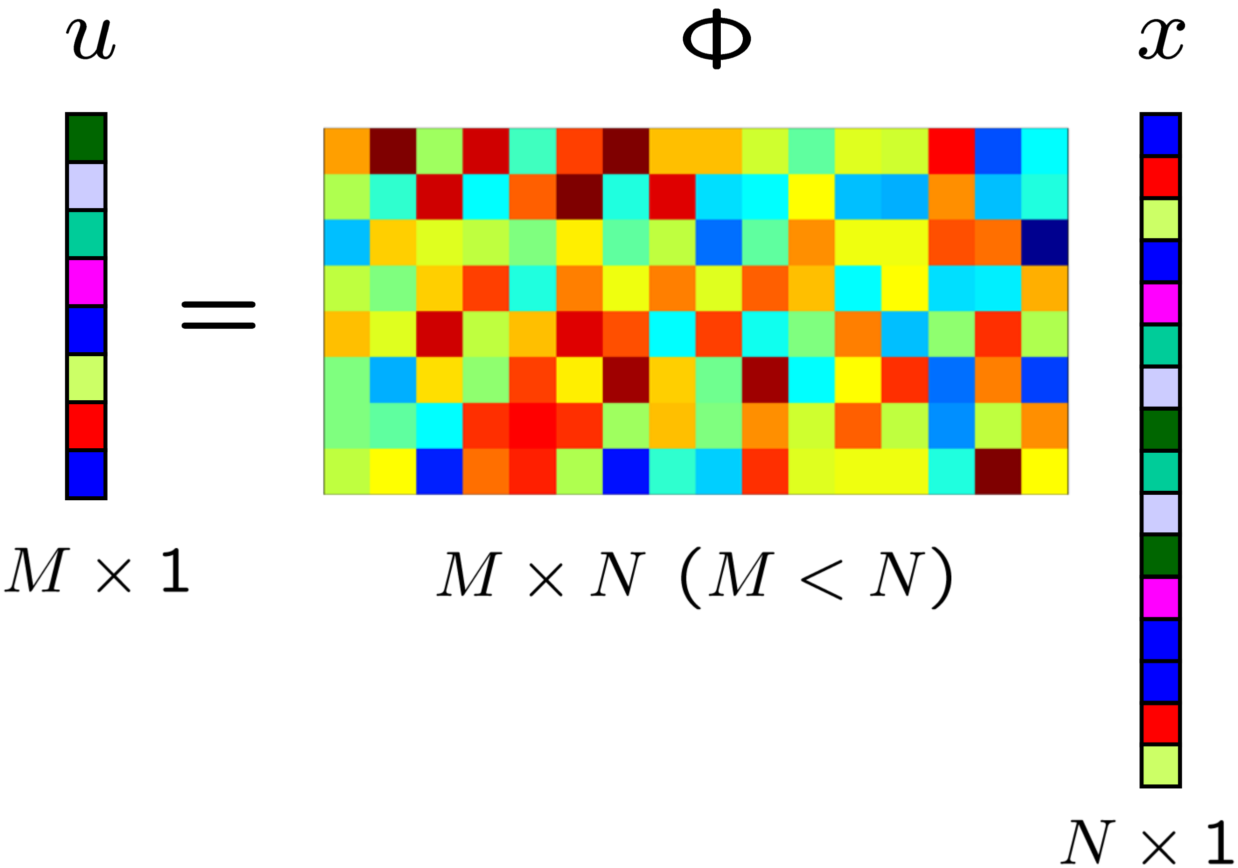
(adversarial) perturbations

observations → $$u = \Phi(X) + n$$

linear (sampling) operator

CS: *when we have less samples than the ambient dimension*

Linear Inverse Problems



Myriad applications involve linear dimensionality reduction

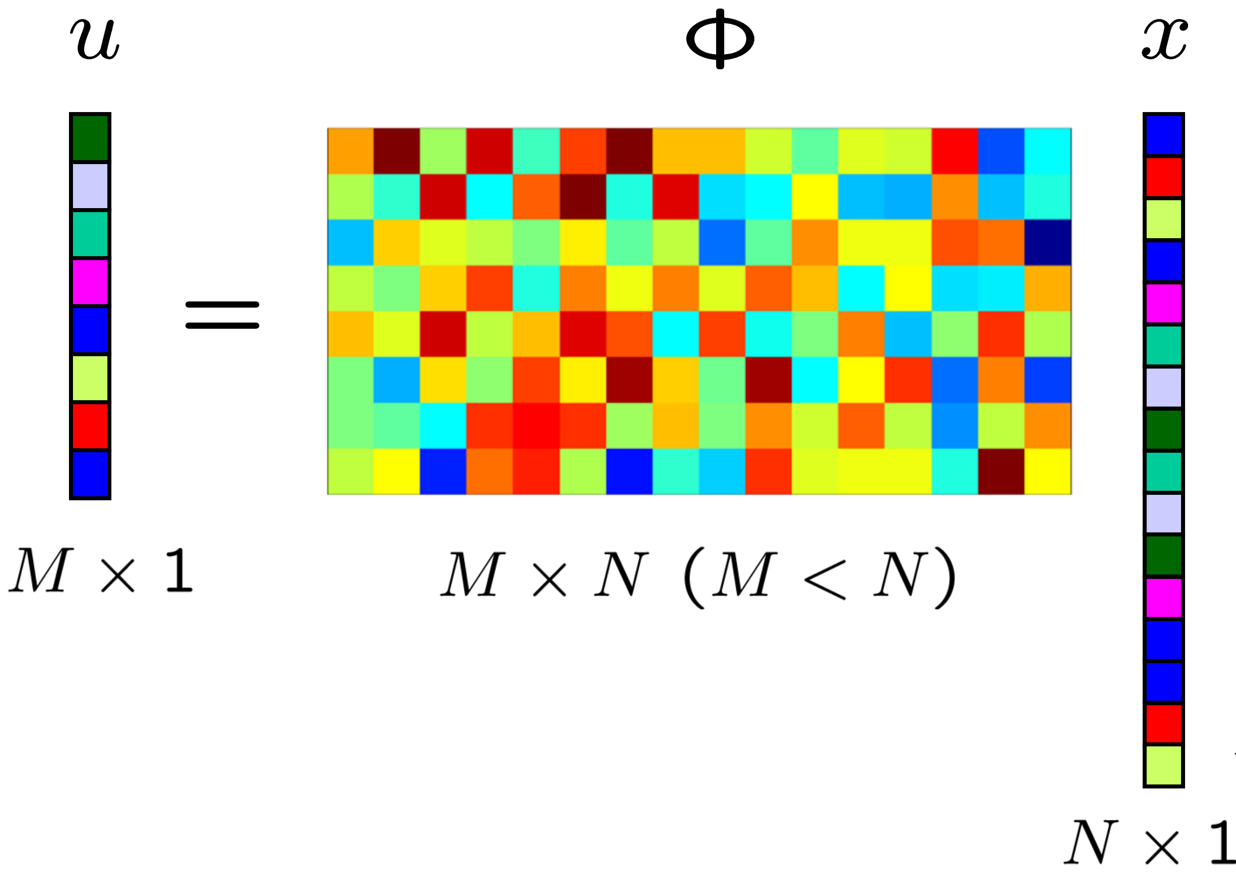
deconvolution to data mining

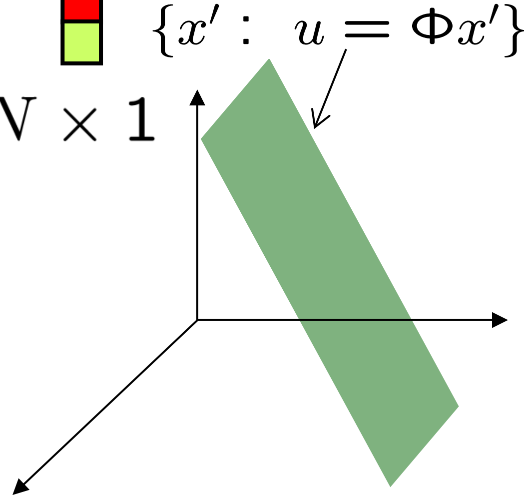
compression to compressive sensing

geophysics to medical imaging

[Baraniuk, C, Wakin 2010; Carin et al. 2011]

Linear Inverse Problems





- **Challenge:** Null space of Φ: $\mathcal{N}(\Phi)$

$$\Phi x' = \Phi(x + v) = u, \;\; \forall v \in \mathcal{N}(\Phi)$$

Linear Inverse Problems

	Deterministic	**Probabilistic**
Prior	Sparsity	distribution
Metric	ℓ_p-norm*	likelihood/ posterior

$$* : \|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$$

Deterministic Low-Dimensional Models

Sparse representations

- **Sparse** signal α

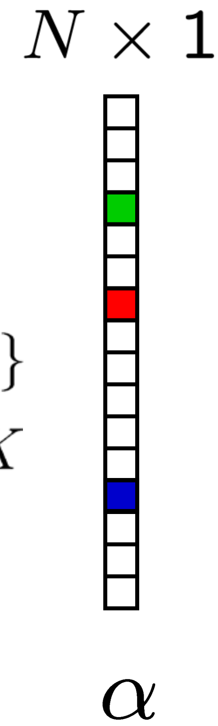
  only K out of N
  coordinates nonzero

  $K \ll N$

support:

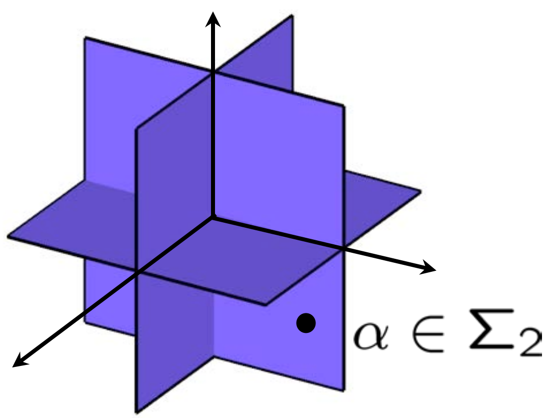
$$\mathcal{S} = \{i : x_i \neq 0\}$$

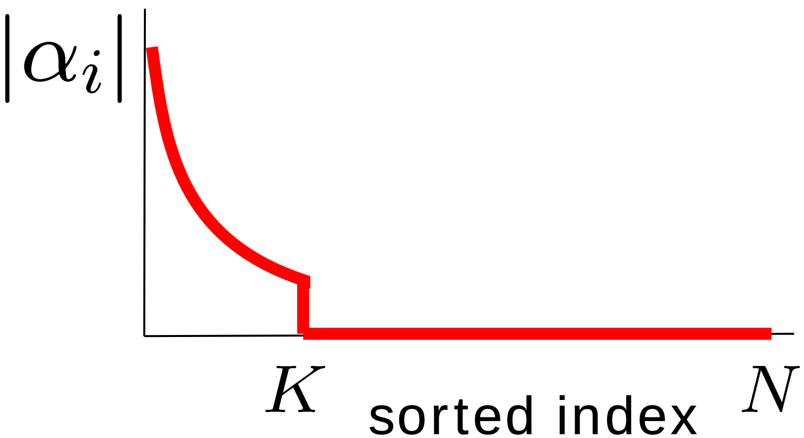
$$\|\alpha\|_0 = |\mathcal{S}| = K$$



$K = 2$

$\mathbf{R}^3$





Sparse representations

- **Sparse** signal x

  only K out of N coordinates nonzero in an ***appropriate representation***

- Sparse representations

  sparse transform coefficients α

  – Basis representations

  $$\Psi \in \mathbb{R}^{N \times N}$$

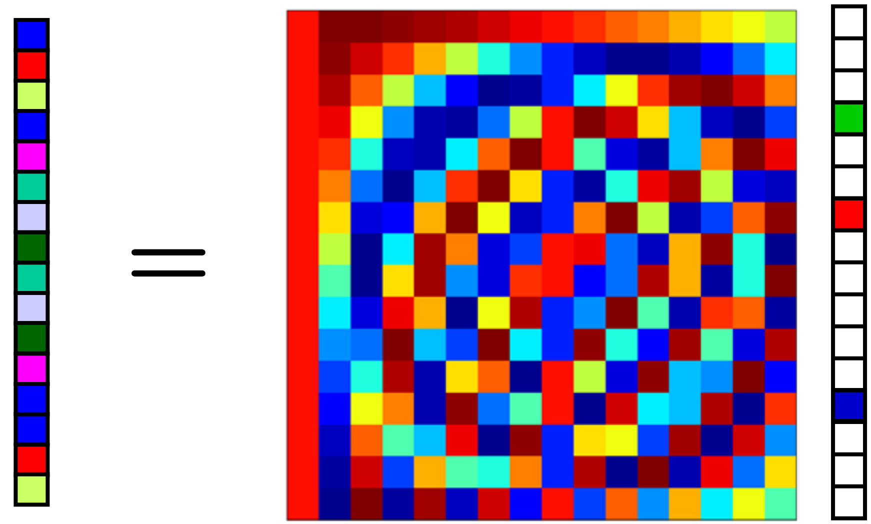
  - **Wavelets**, DCT…

  – Frame representations

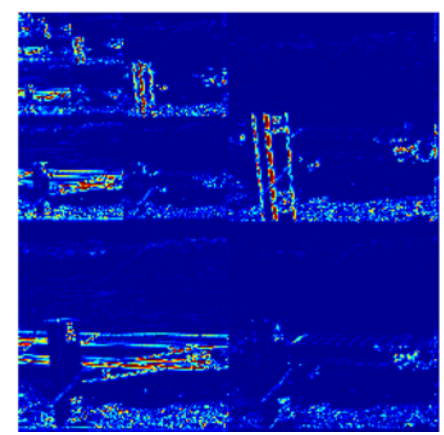
  $$\Psi \in \mathbb{R}^{N \times L}, L > N$$

  - Gabor, curvelets, shearlets…

  – Other ***dictionary*** representations…



$$x = \Psi \times \alpha$$



Sparse representations

- Sparse signal:

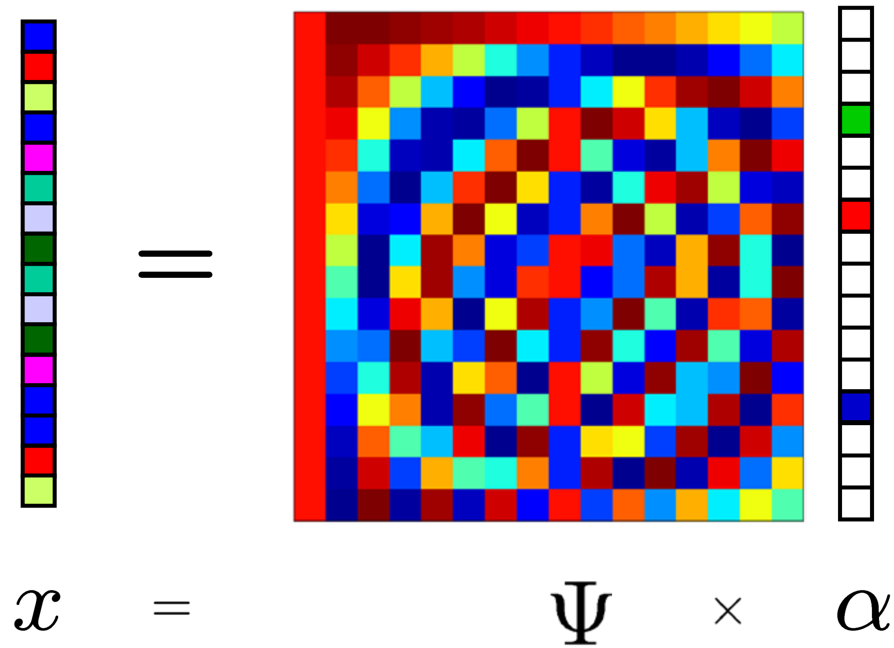
  only K out of N
  coordinates nonzero

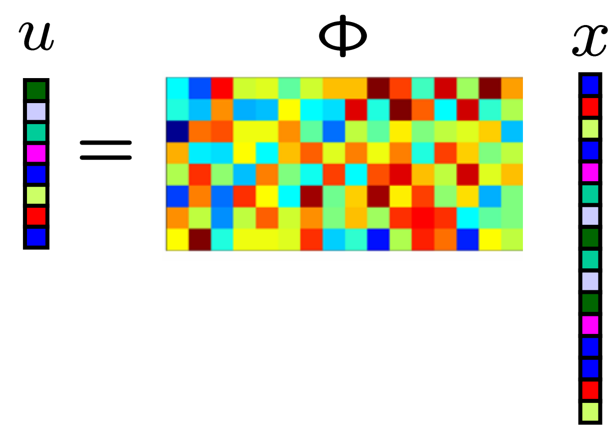
  $$K \ll N$$

- Sparse representations:

  sparse transform
  coefficients

- A fundamental impact:





Sparse representations

- Sparse signal:

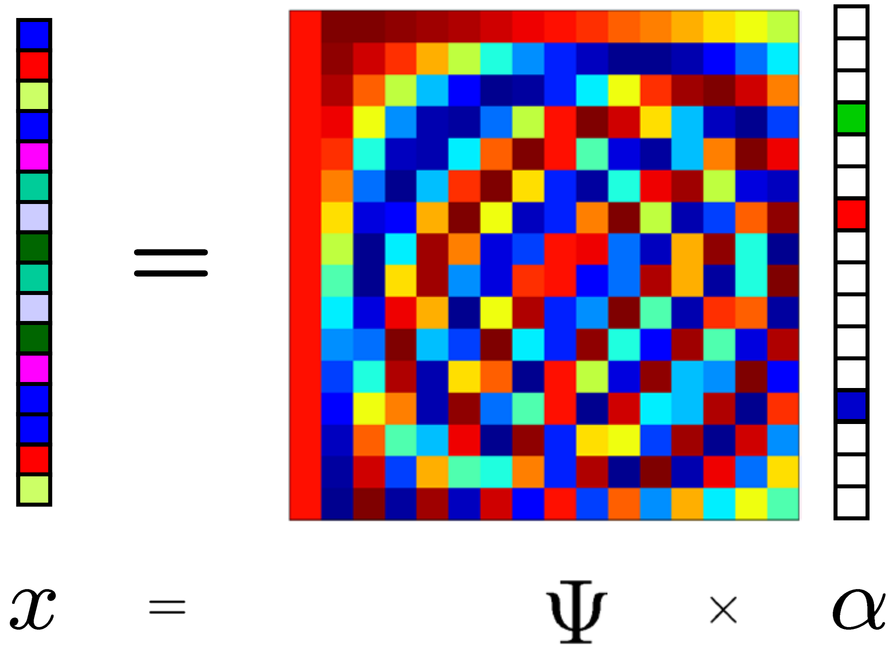
  only K out of N
  coordinates nonzero

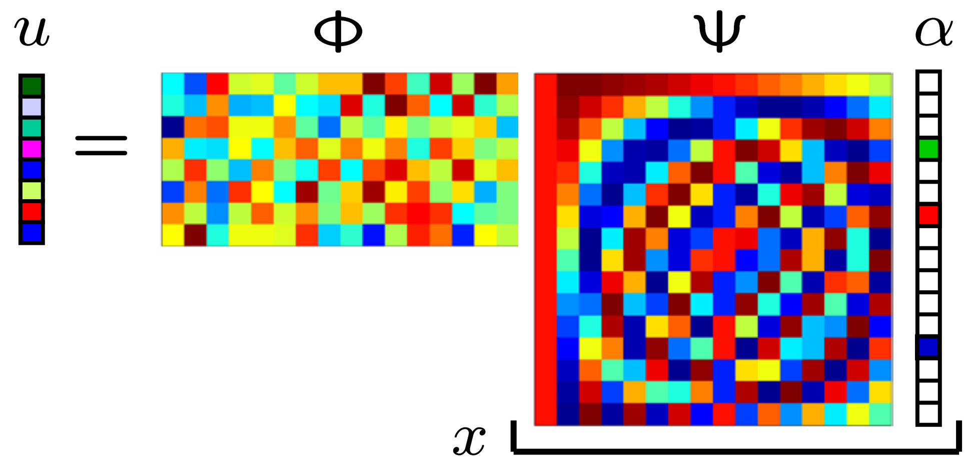
  $$K \ll N$$

- Sparse representations:

  sparse transform
  coefficients

- A fundamental impact:



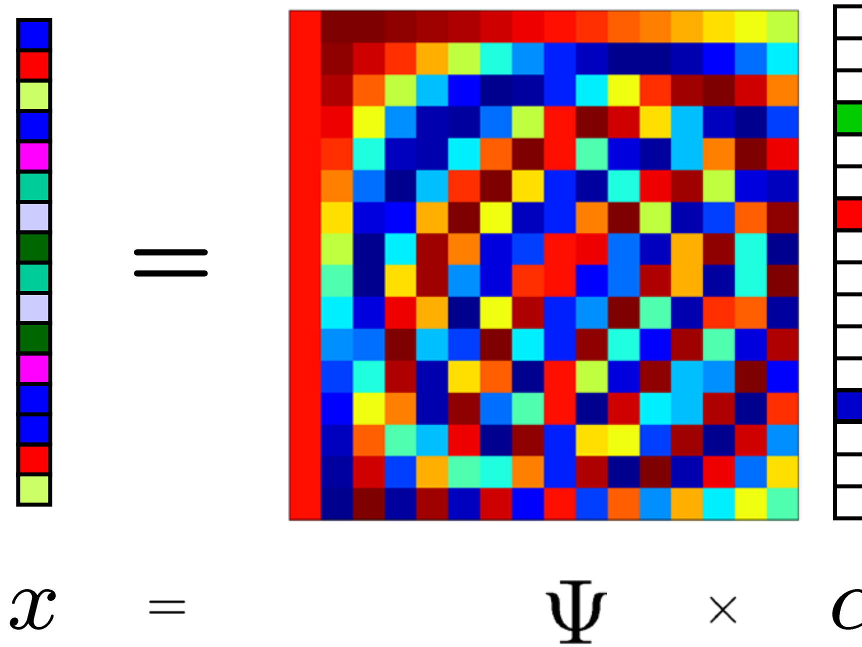


Sparse representations

- Sparse signal:

  only K out of N coordinates nonzero

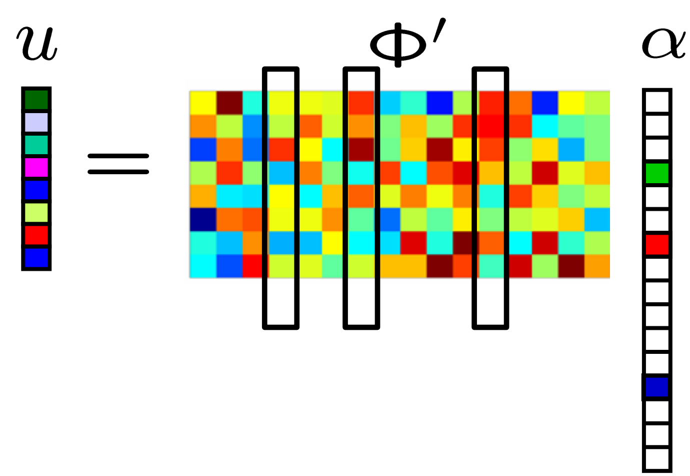
  $$K \ll N$$



- Sparse representations:

  sparse transform coefficients

- A fundamental impact:



Sparse representations

- Sparse signal:

  only K out of N coordinates nonzero

  $$K \ll N$$

- Sparse representations:

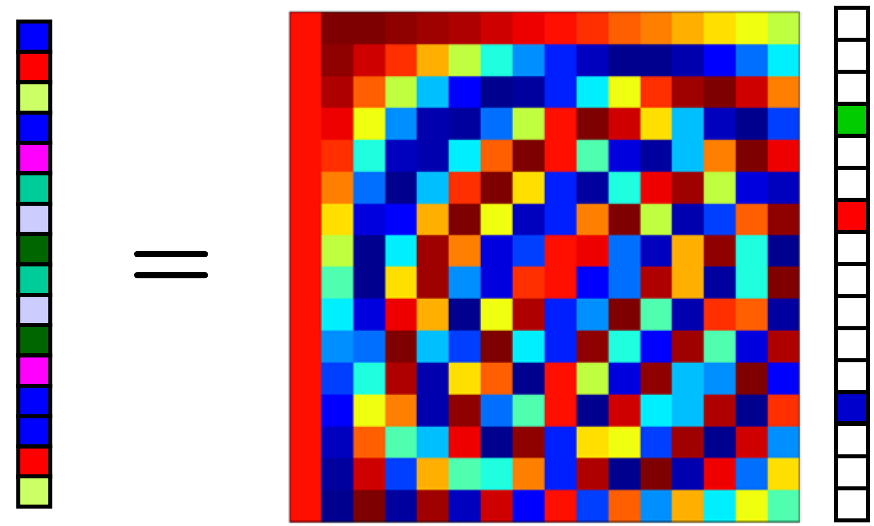
  sparse transform coefficients

- A fundamental impact:

  $$\Phi$$

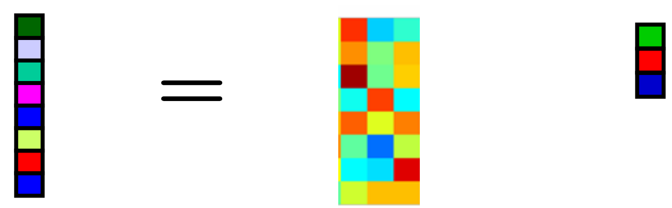
  becomes effectively low dimensional*

  $$M \times K$$



$$x = \Psi \times \alpha$$

$$u \quad \Phi' \quad \alpha$$



$$M > K$$

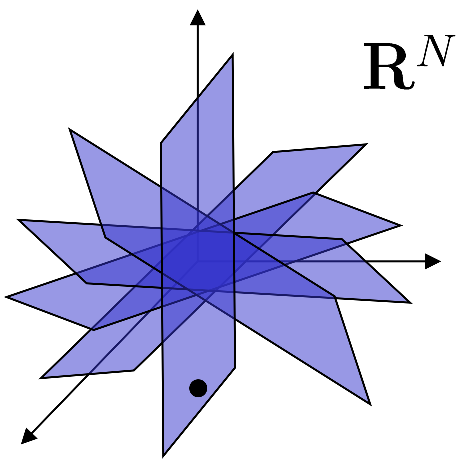
*: If we knew the locations of the coefficients. **More on this later.**

Low-dimensional signal models

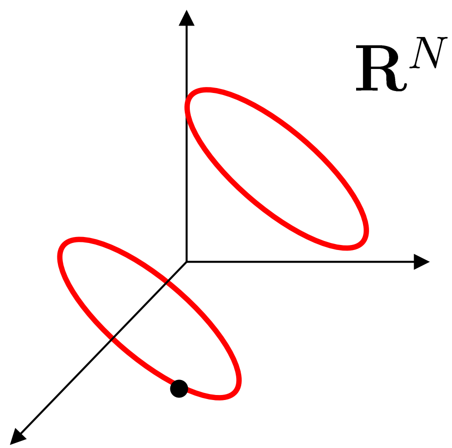
N
pixels

Information level:

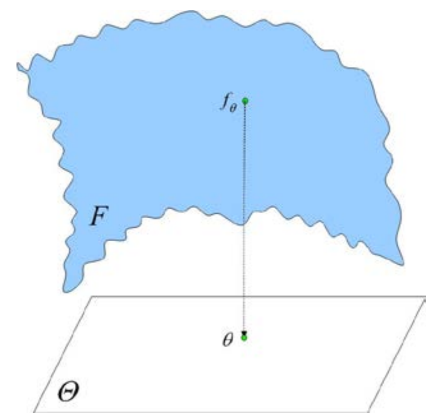
$K \ll N$
large
wavelet
coefficients
(blue = 0)



sparse
signals



low-rank
matrices

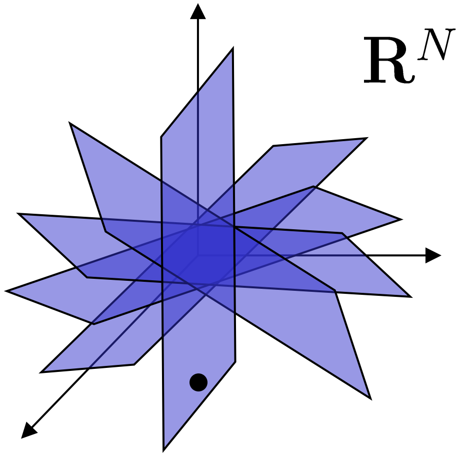


nonlinear
models

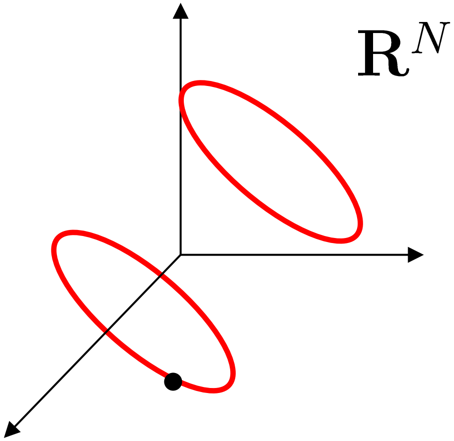
Low-dimensional signal models

- These lectures

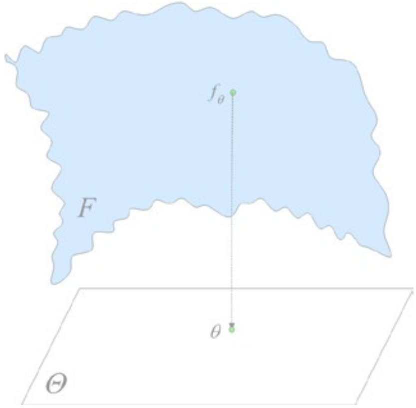
Low-dimensional models
based on
linear representations



sparse
signals



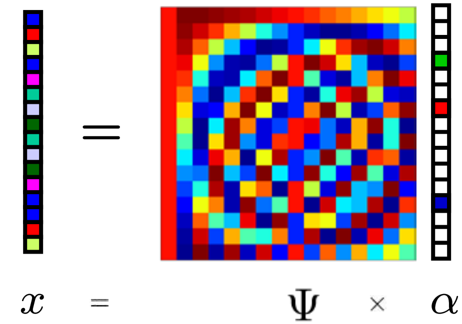
low-rank
matrices



nonlinear
models

Linear representation of low-dimensional models

- A key notion in sparse representation
  - synthesis of the signal using a few vectors



- A slightly different mathematical formalism for generalization

Synthesis model:

$$x = \sum_{i=1}^{|\mathcal{A}|} a_i c_i \qquad a_i \in \mathcal{A}, c_i \geq 0$$

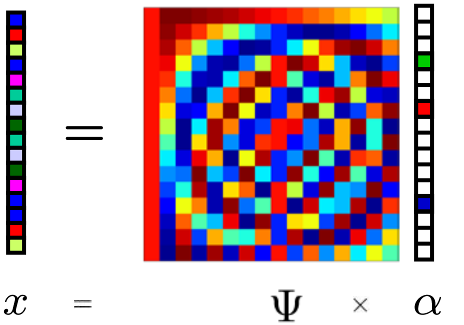
a_i: atoms
$\mathcal{A}$: atomic set

i.e., linear (positive) combination of elements from an atomic set

[Chandrasekaran et al. 2010]

Linear representation of low-dimensional models

- A key notion in sparse representation



  - synthesis of the signal using a few vectors

- Sparse representations via the atomic formulation

$$x = \sum_{i=1}^{|\mathcal{A}|} a_i c_i$$

$a_i \in \mathcal{A}, c_i \geq 0$

a_i: atoms

$\mathcal{A}$: atomic set

  - Example:

$$\Psi = [\psi_1, \ldots, \psi_L]$$

$$\text{rank}(\Psi) = N$$

$$\mathcal{A} = \{\psi_1, ..., \psi_L, -\psi_1, \ldots, -\psi_L\}$$

$$c_i = \begin{cases} \alpha_i, & \alpha_i > 0; \\ 0, & \text{otherwise.} \end{cases} \qquad i = 1, \ldots, L$$

$$c_{i+L} = \begin{cases} -\alpha_i, & \alpha_i < 0; \\ 0, & \text{otherwise.} \end{cases}$$

Linear representation of low-dimensional models

- Basic definitions on **low-dimensional** *atomic representations*

$$x = \sum_{i=1}^{|\mathcal{A}|} a_i c_i \qquad a_i \in \mathcal{A}, c_i \geq 0 \qquad K \ll N$$

$$\|c_i\|_0 \leq K$$

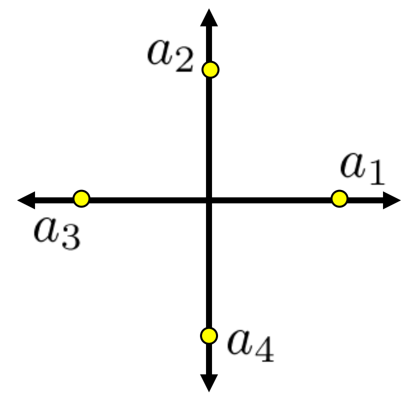
Linear representation of low-dimensional models

- Basic definitions on low-dimensional *atomic representations*

$$x = \sum_{i=1}^{|\mathcal{A}|} a_i c_i \qquad a_i \in \mathcal{A}, c_i \geq 0 \qquad \|c_i\|_0 \leq K \qquad K \ll N$$

  - $\mathrm{conv}(\mathcal{A})$: convex hull of atoms in A

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$$



$$\mathrm{conv}(\mathcal{A}) = \{\textstyle\sum_i a_i \beta_i : a_i \in \mathcal{A}, \beta_i \in \mathbb{R}_+, \sum_{i=1}^n \beta_i = 1, \; n = 1, 2, \ldots, |\mathcal{A}|\}$$

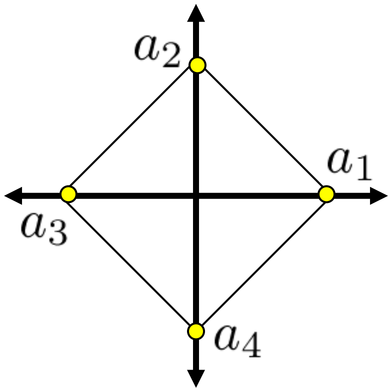
Linear representation of low-dimensional models

- Basic definitions on low-dimensional *atomic representations*

$$x = \sum_{i=1}^{|\mathcal{A}|} a_i c_i \qquad a_i \in \mathcal{A}, c_i \geq 0 \qquad \|c_i\|_0 \leq K \qquad K \ll N$$

  - $\text{conv}(\mathcal{A})$: convex hull of atoms in A

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$$



atomic ball

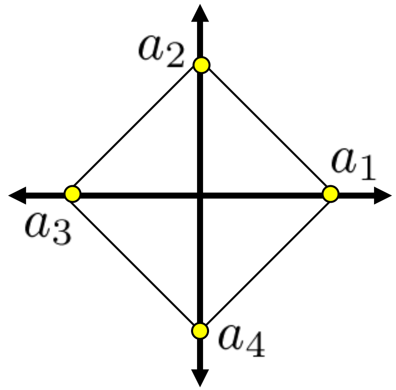
$$\text{conv}(\mathcal{A}) = \{\textstyle\sum_i a_i \beta_i : a_i \in \mathcal{A}, \beta_i \in \mathbb{R}_+, \sum_{i=1}^{n} \beta_i = 1, \, n = 1, 2, \ldots, |\mathcal{A}|\}$$

Linear representation of low-dimensional models

- Basic definitions on low-dimensional *atomic representations*

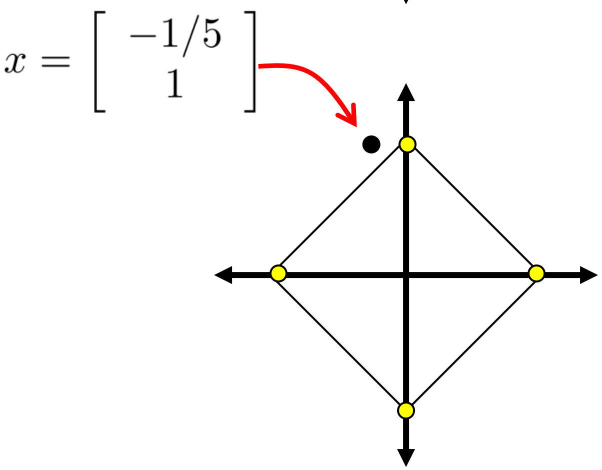
$$x = \sum_{i=1}^{|\mathcal{A}|} a_i c_i \qquad a_i \in \mathcal{A}, c_i \geq 0 \qquad \|c_i\|_0 \leq K \qquad K \ll N$$

  - $\mathrm{conv}(\mathcal{A})$: convex hull of atoms in A



$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$$

  - $\|x\|_{\mathcal{A}}$: atomic norm*



$$\|x\|_{\mathcal{A}} = \inf\{t > 0 : x \in t \times \mathrm{conv}(\mathcal{A})\}$$

*: requires A to be centrally symmetric

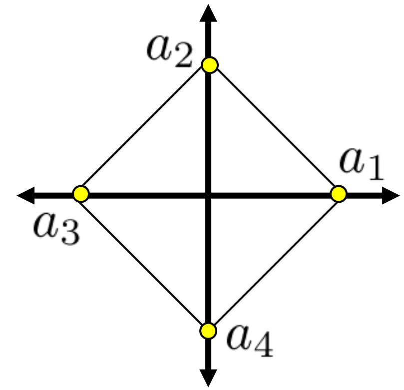
Linear representation of low-dimensional models

- Basic definitions on low-dimensional *atomic representations*

$$x = \sum_{i=1}^{|\mathcal{A}|} a_i c_i \qquad a_i \in \mathcal{A}, c_i \geq 0 \qquad \|c_i\|_0 \leq K \qquad K \ll N$$

  - $\text{conv}(\mathcal{A})$: convex hull of atoms in A

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$$

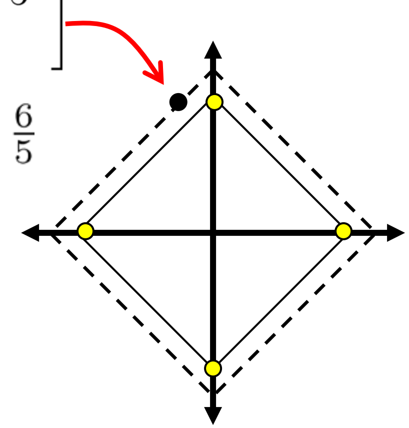


  - $\|x\|_{\mathcal{A}}$: atomic norm*

$$\|x\|_{\mathcal{A}} = \inf\{t > 0 : x \in t \times \text{conv}(\mathcal{A})\}$$

$$x = \begin{bmatrix} -1/5 \\ 1 \end{bmatrix}$$

$$\|x\|_{\mathcal{A}} = \frac{6}{5}$$



*: requires A to be centrally symmetric

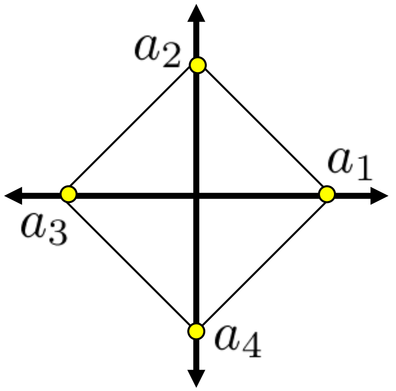
Linear representation of low-dimensional models

- Basic definitions on low-dimensional *atomic representations*

$$x = \sum_{i=1}^{|\mathcal{A}|} a_i c_i \qquad a_i \in \mathcal{A}, c_i \geq 0 \qquad \|c_i\|_0 \leq K \qquad K \ll N$$

  - $\mathrm{conv}(\mathcal{A})$: convex hull of atoms in A

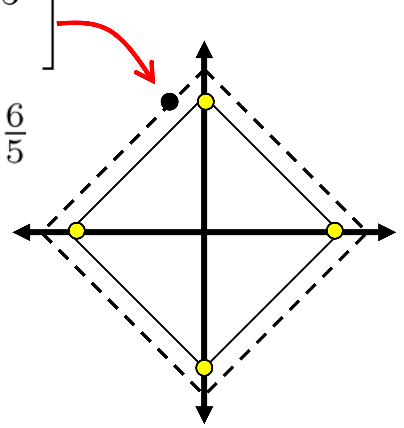
$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$$



  - $\|x\|_{\mathcal{A}}$: atomic norm*

$$\|x\|_{\mathcal{A}} = \inf\{t > 0 : x \in t \times \mathrm{conv}(\mathcal{A})\}$$

$$x = \begin{bmatrix} -1/5 \\ 1 \end{bmatrix} \qquad \|x\|_{\mathcal{A}} = \frac{6}{5}$$



Alternative: $\|x\|_{\mathcal{A}} = \inf\left\{ \sum_{i=1}^{|\mathcal{A}|} c_i : x = \sum_{i=1}^{|\mathcal{A}|} a_i c_i, c_i \geq 0, \forall a_i \in \mathcal{A} \right\}$

*: requires A to be centrally symmetric

Linear representation of low-dimensional models

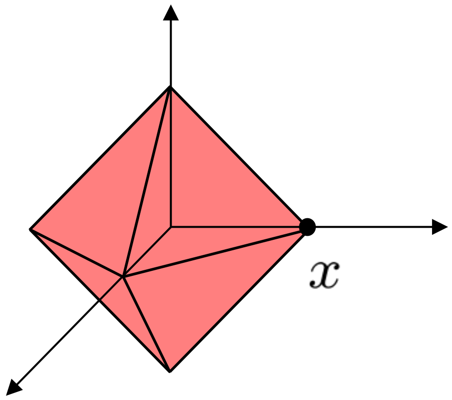
Examples with easy forms:

- *sparse vectors*

$$\mathcal{A} = \{\pm e_i\}_{i=1}^{N}$$

$$\text{conv}(\mathcal{A}) = \text{cross-polytope}$$

$$\|x\|_{\mathcal{A}} = \|x\|_1$$

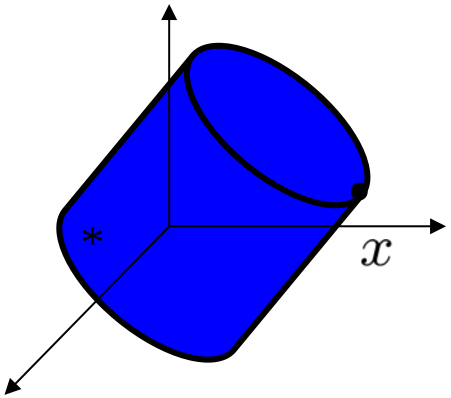


- *low-rank matrices*

$$\mathcal{A} = \{A : \text{rank}(A) = 1, \|A\|_F = 1\}$$

$$\text{conv}(\mathcal{A}) = \text{nuclear norm ball}$$

$$\|x\|_{\mathcal{A}} = \|x\|_\star$$



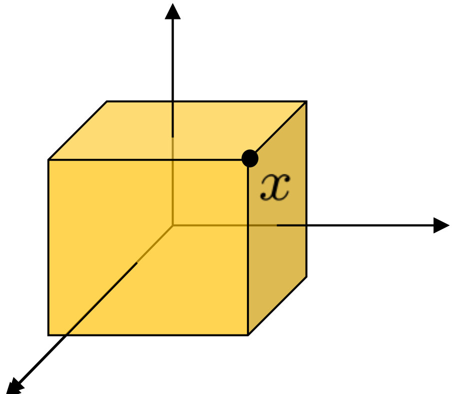
*symmetric matrices

- *binary vectors*

$$\mathcal{A} = \{\pm 1\}^N$$

$$\text{conv}(\mathcal{A}) = \text{hypercube}$$

$$\|x\|_{\mathcal{A}} = \|x\|_\infty$$



Linear representation of low-dimensional models

Examples with easy forms:

- *sparse vectors*

$\mathcal{A} = \{\pm e_i\}_{i=1}^{N}$

Examples with no-so-easy forms:

✓**A : infinite set of unit-norm rank-one tensors**

✓**A : finite (but large) set of permutation matrices**

✓**A : infinite set of orthogonal matrices**

✓**A : infinite set of matrices constrained by eigenvalues**

✓**A : infinite set of measures**

✓**A : finite (but large) set of cut matrices**

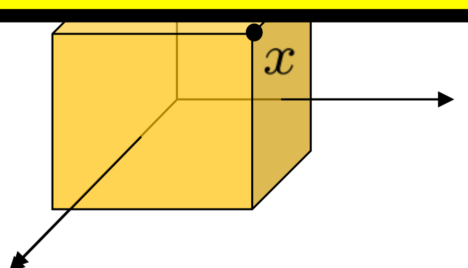
[Chandrasekaran et al. 2010]

- *l*

- *b*

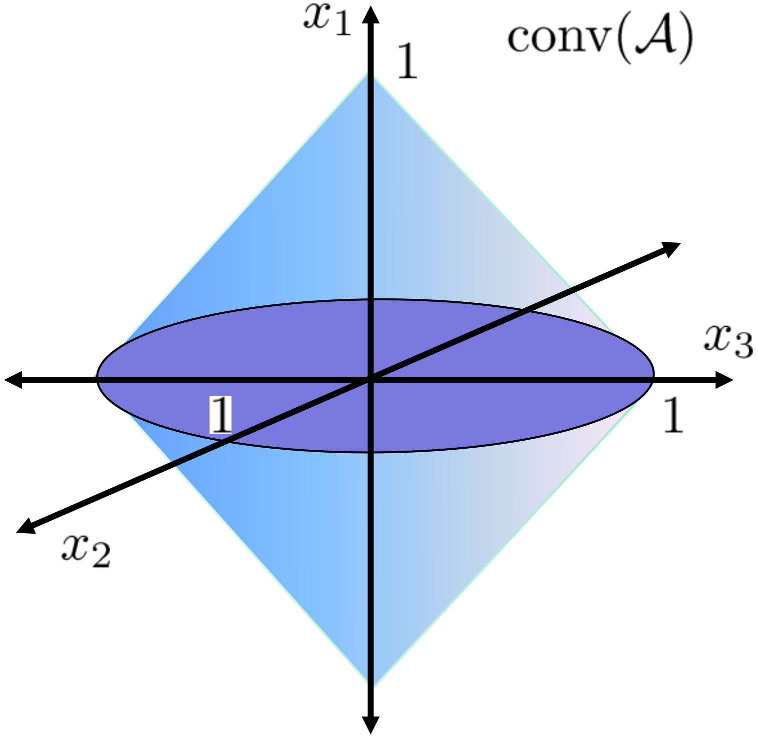
$\mathrm{conv}(\mathcal{A}) = \mathrm{hypercube}$

$\|x\|_{\mathcal{A}} = \|x\|_\infty$



Linear representation of low-dimensional models

Pop-quiz:

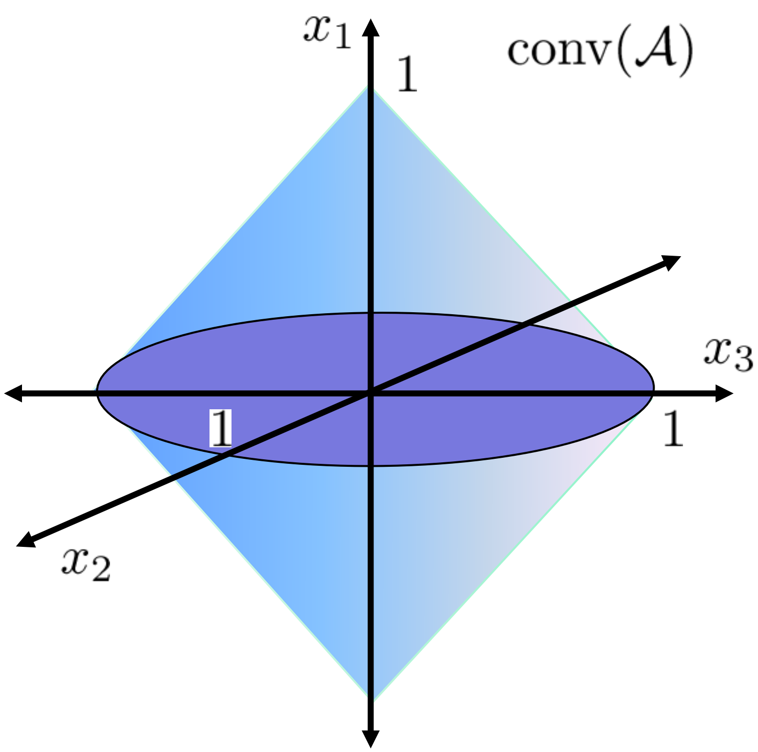


What is $\|x\|_{\mathcal{A}}$ **?**

$$\|x\|_{\mathcal{A}} = \inf\{t > 0 : x \in t \times \text{conv}(\mathcal{A})\}$$

Linear representation of low-dimensional models

Pop-quiz:



HINT:

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \|x_G\|_2 = 1 \right\}$$

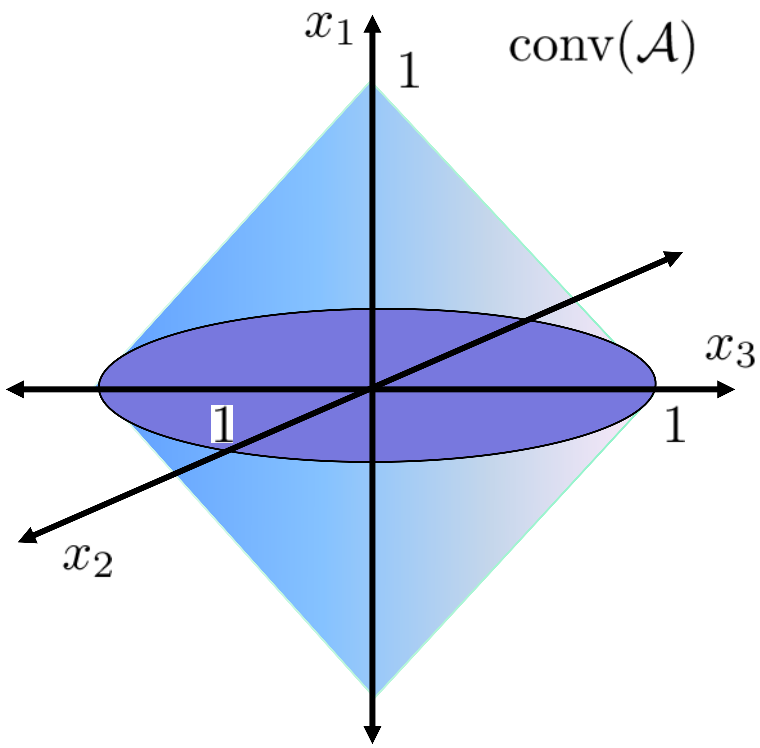
$$G = \{2, 3\}$$

What is $\|x\|_{\mathcal{A}}$ **?**

$$\|x\|_{\mathcal{A}} = \inf\{t > 0 : x \in t \times \text{conv}(\mathcal{A})\}$$

Linear representation of low-dimensional models

Pop-answer:



$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \|x_G\|_2 = 1 \right\}$$

What is $\|x\|_{\mathcal{A}}$ **?**

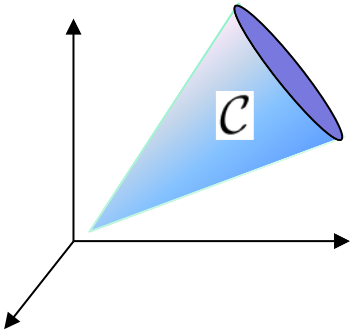
$$\|x\|_{\mathcal{A}} = |x_1| + \|x_G\|_2$$

$$G = \{2, 3\}$$

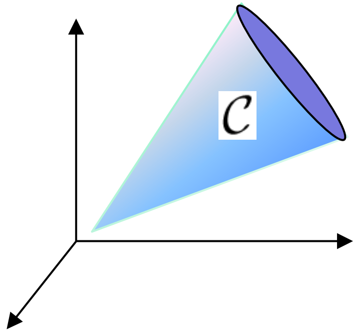
Towards algorithms: a geometric perspective

Other key concepts:

- Cone $\mathcal{C}$: $x, y \in \mathcal{C} \Rightarrow tx + \omega y \in \mathcal{C}, \forall t, \omega \in \mathbb{R}_+$



Towards algorithms: a geometric perspective

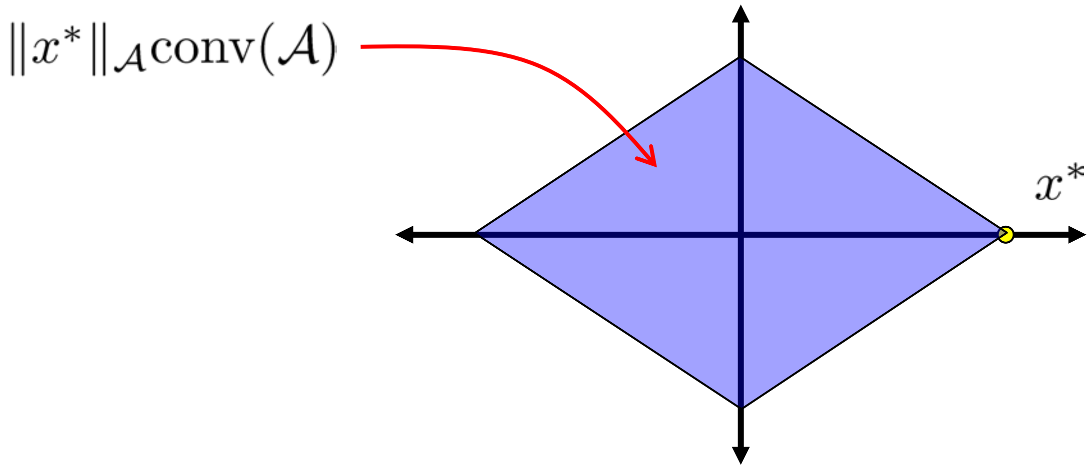


Other key concepts:

- Cone $\mathcal{C}$: $x, y \in \mathcal{C} \Rightarrow tx + \omega y \in \mathcal{C}, \forall t, \omega \in \mathbb{R}_+$

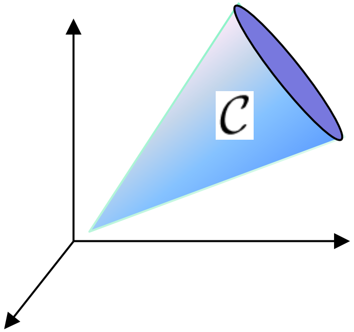
- Tangent cone of x^* with respect to $\|x^*\|_{\mathcal{A}}\text{conv}(\mathcal{A})$:

$$T_{\mathcal{A}}(x^*) = \text{cone}\{z - x^* : \|z\|_{\mathcal{A}} \leq \|x^*\|_{\mathcal{A}}\}$$



Towards algorithms: a geometric perspective

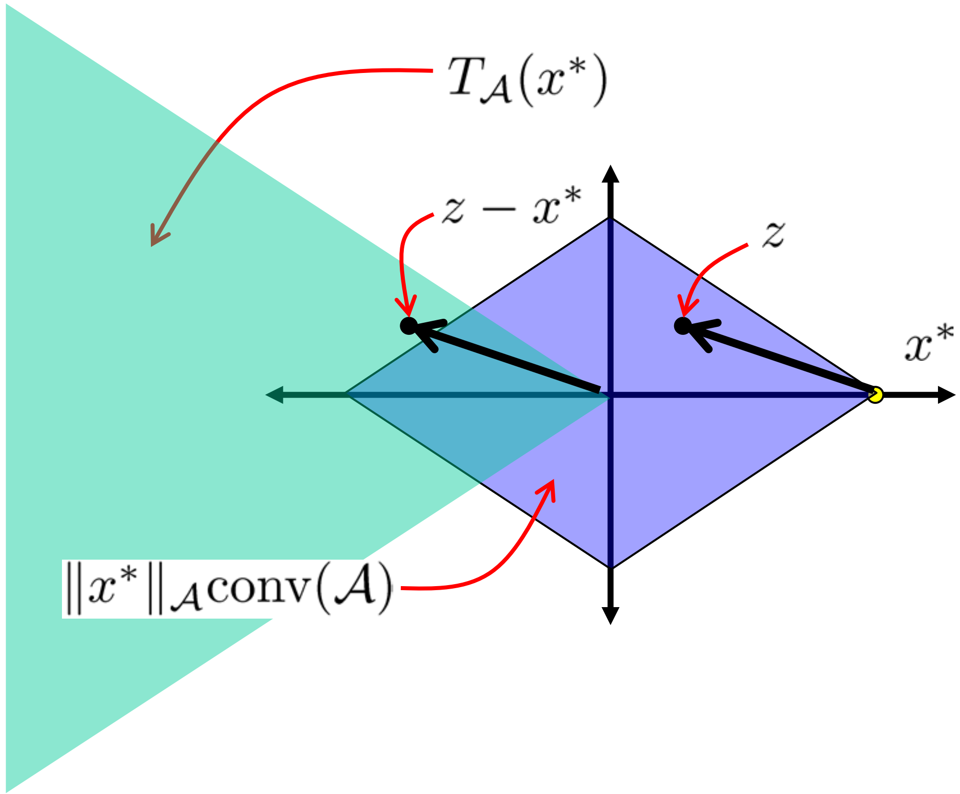
Other key concepts:



- Cone $\mathcal{C}$: $x, y \in \mathcal{C} \Rightarrow tx + \omega y \in \mathcal{C}, \forall t, \omega \in \mathbb{R}_+$

- Tangent cone of x^* with respect to $\|x^*\|_{\mathcal{A}}\text{conv}(\mathcal{A})$:

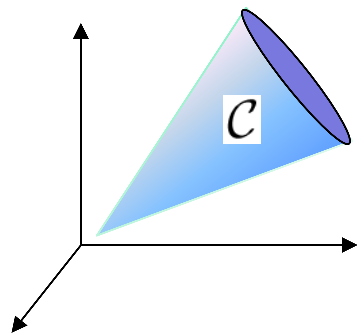
$$T_{\mathcal{A}}(x^*) = \text{cone}\{z - x^* : \|z\|_{\mathcal{A}} \le \|x^*\|_{\mathcal{A}}\}$$



Tangent cone
is the set of descent directions where you do not increase the atomic norm.

Towards algorithms: a geometric perspective

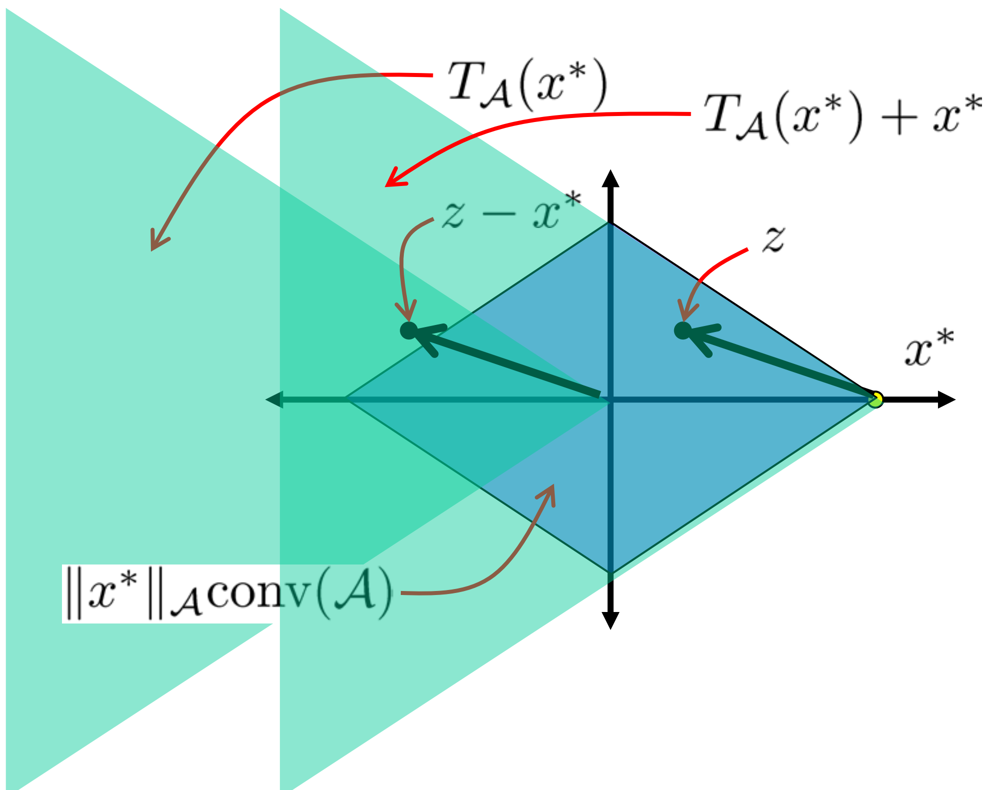
Other key concepts:



- Cone $\mathcal{C}$: $x, y \in \mathcal{C} \Rightarrow tx + \omega y \in \mathcal{C}, \forall t, \omega \in \mathbb{R}_+$

- Tangent cone of x^* with respect to $\|x^*\|_{\mathcal{A}}\text{conv}(\mathcal{A})$:

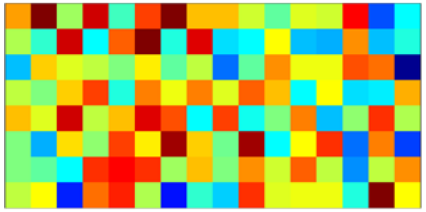
$$T_{\mathcal{A}}(x^*) = \text{cone}\{z - x^* : \|z\|_{\mathcal{A}} \leq \|x^*\|_{\mathcal{A}}\}$$



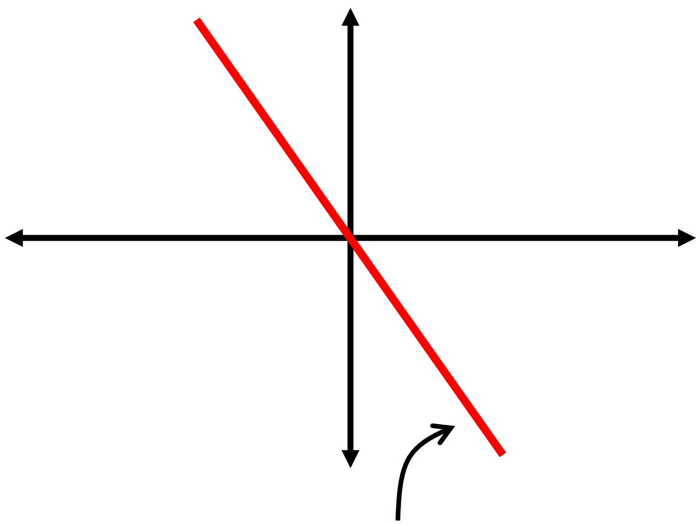
Tangent cone
is the set of descent directions where you do not increase the atomic norm.

Towards algorithms: a geometric perspective

Φ



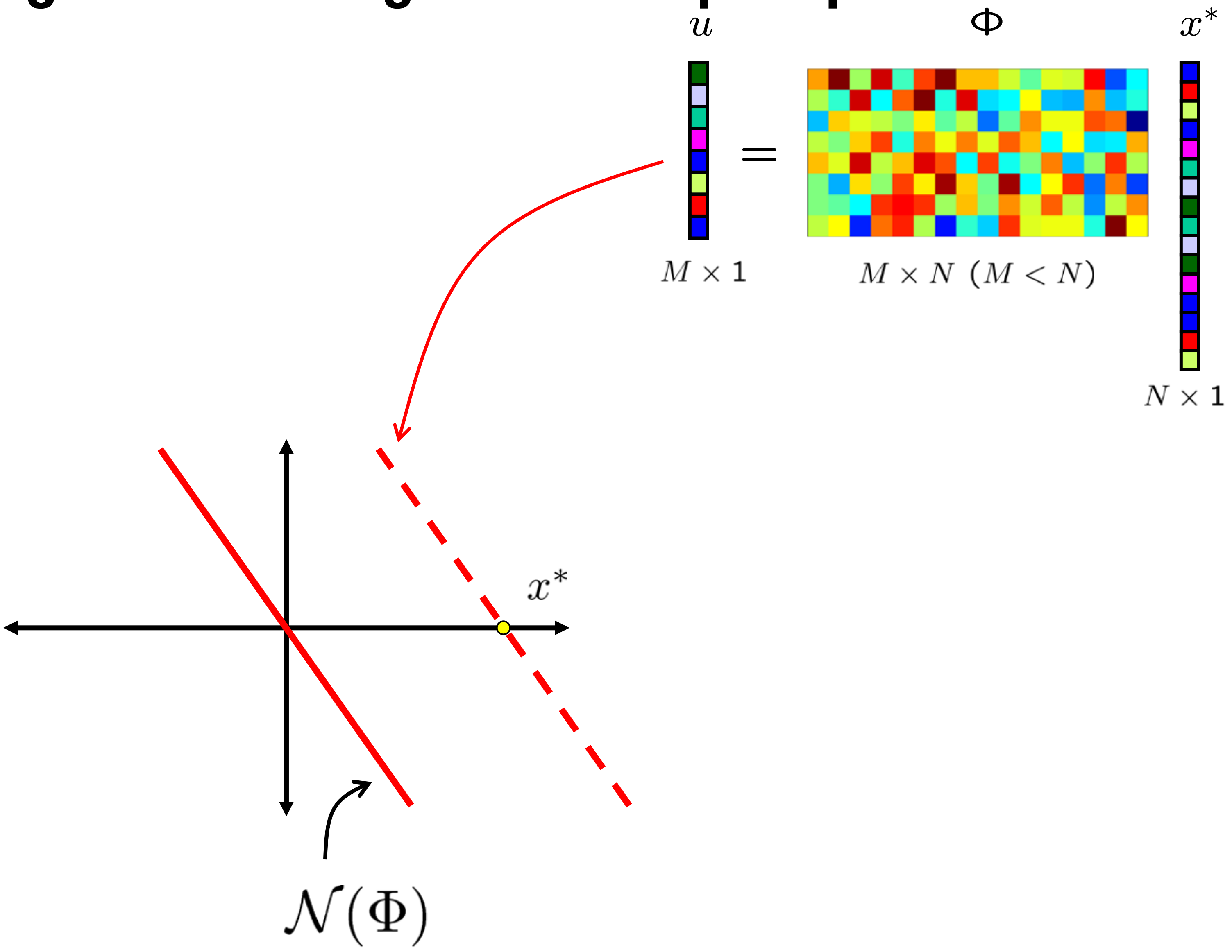
$M \times N \ (M < N)$



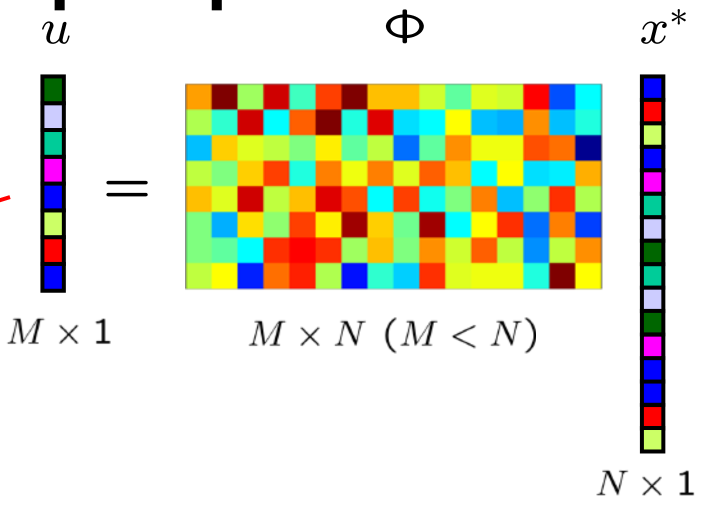
Null space of Φ: $\mathcal{N}(\Phi)$

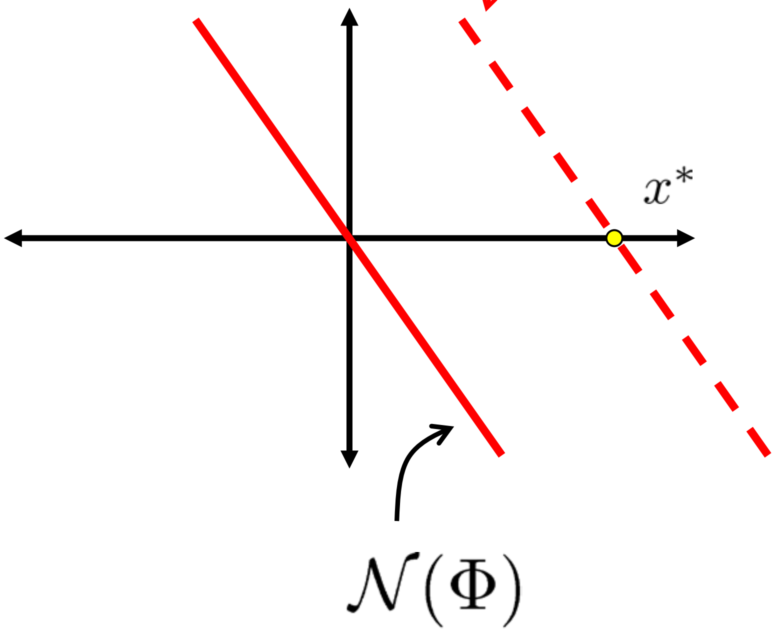
$$\Phi v = 0, \ \forall v \in \mathcal{N}(\Phi)$$

Towards algorithms: a geometric perspective



Towards algorithms: a geometric perspective

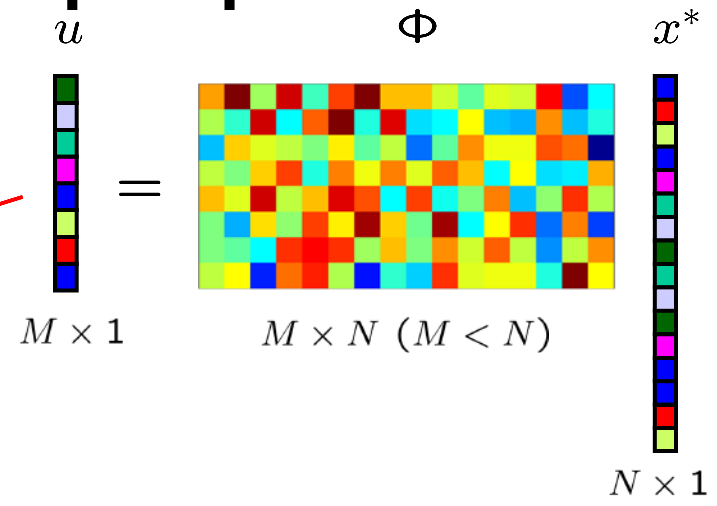


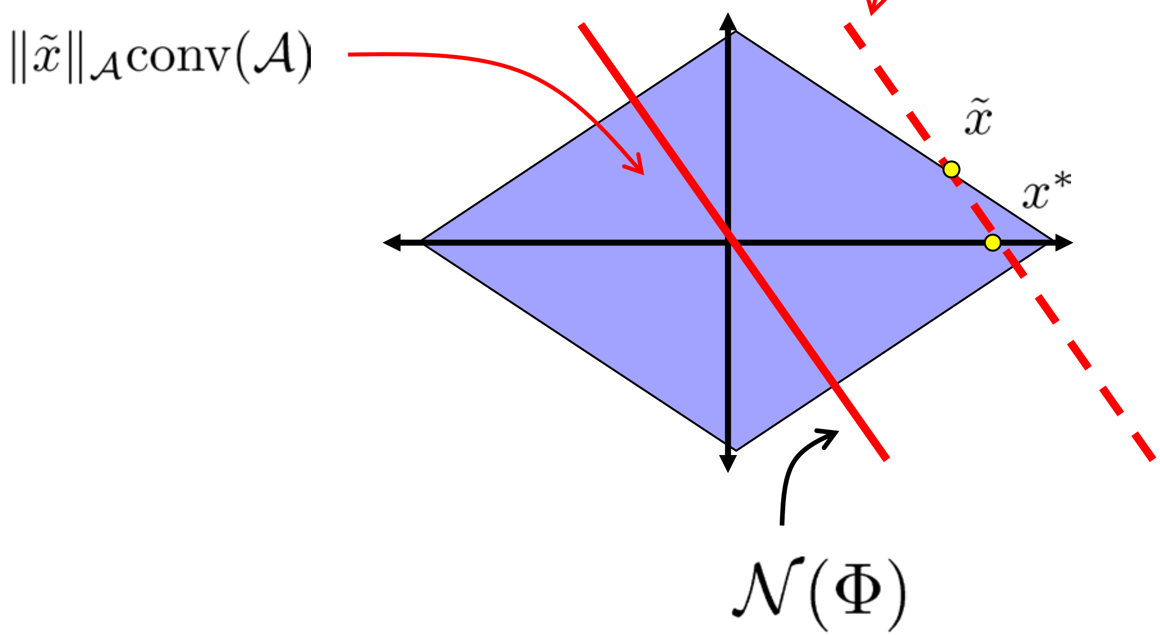


Consider the criteria:

$$\widehat{x} = \arg \min_{x : u = \Phi x} \|x\|_{\mathcal{A}}$$

Towards algorithms: a geometric perspective

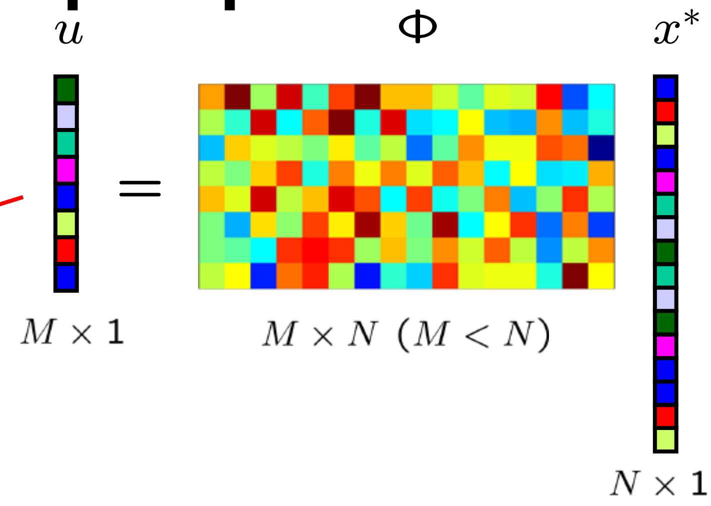


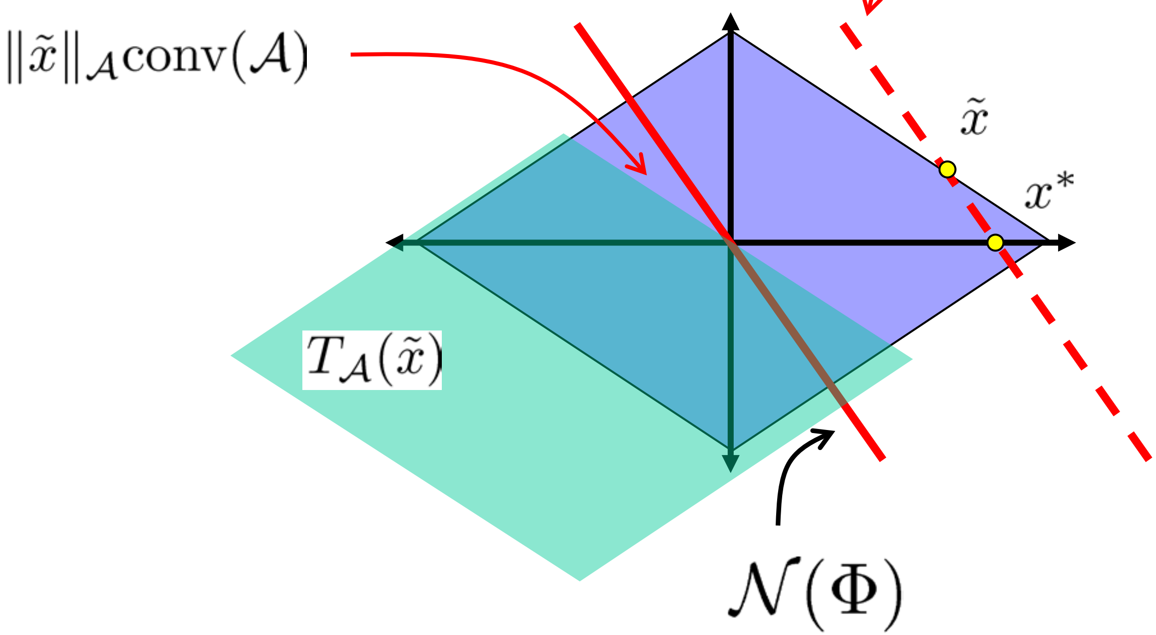


Consider the criteria:

$$\widehat{x} = \arg \min_{x : u = \Phi x} \|x\|_{\mathcal{A}}$$

Towards algorithms: a geometric perspective

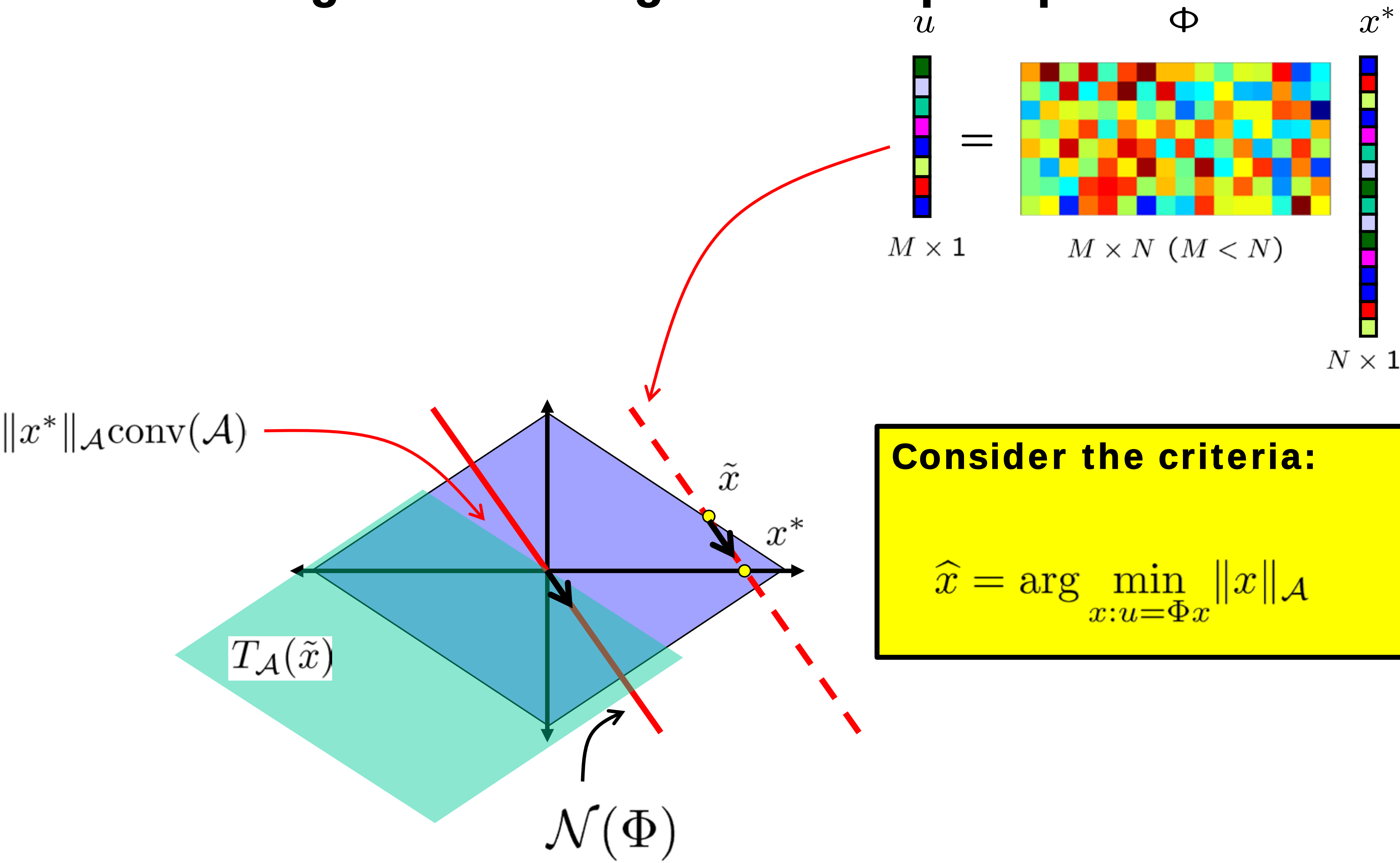




Consider the criteria:

$$\widehat{x} = \arg \min_{x: u = \Phi x} \|x\|_{\mathcal{A}}$$

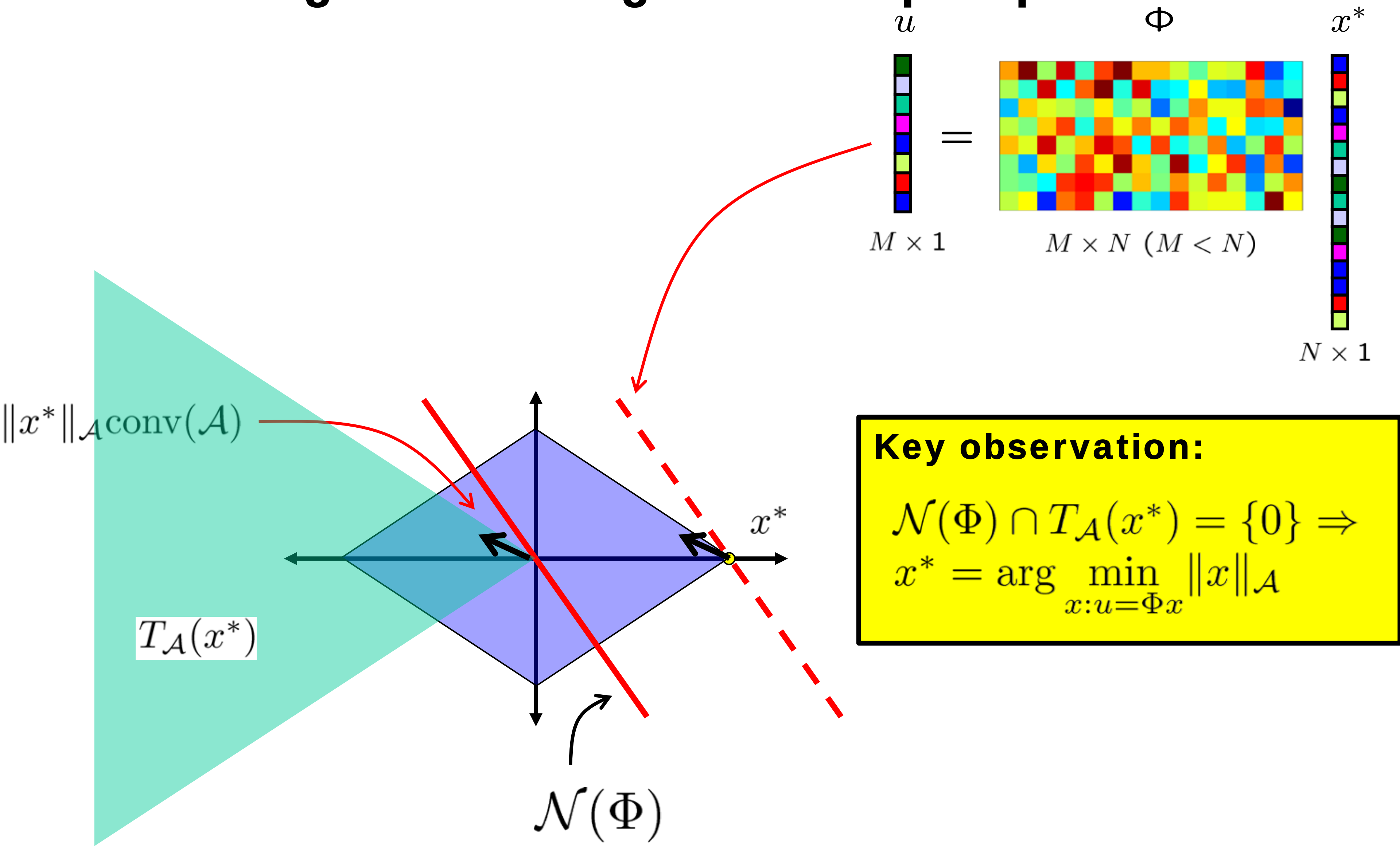
Towards algorithms: a geometric perspective



Consider the criteria:

$$\widehat{x} = \arg \min_{x : u = \Phi x} \|x\|_{\mathcal{A}}$$

Towards algorithms: a geometric perspective

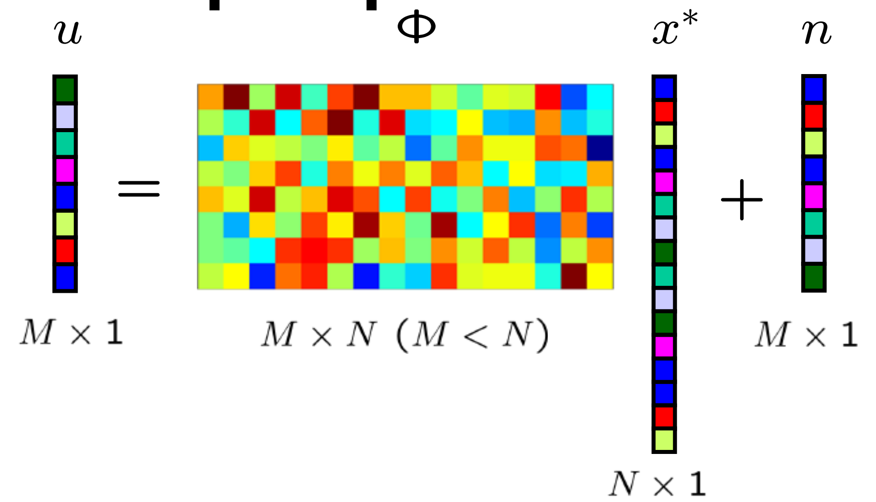


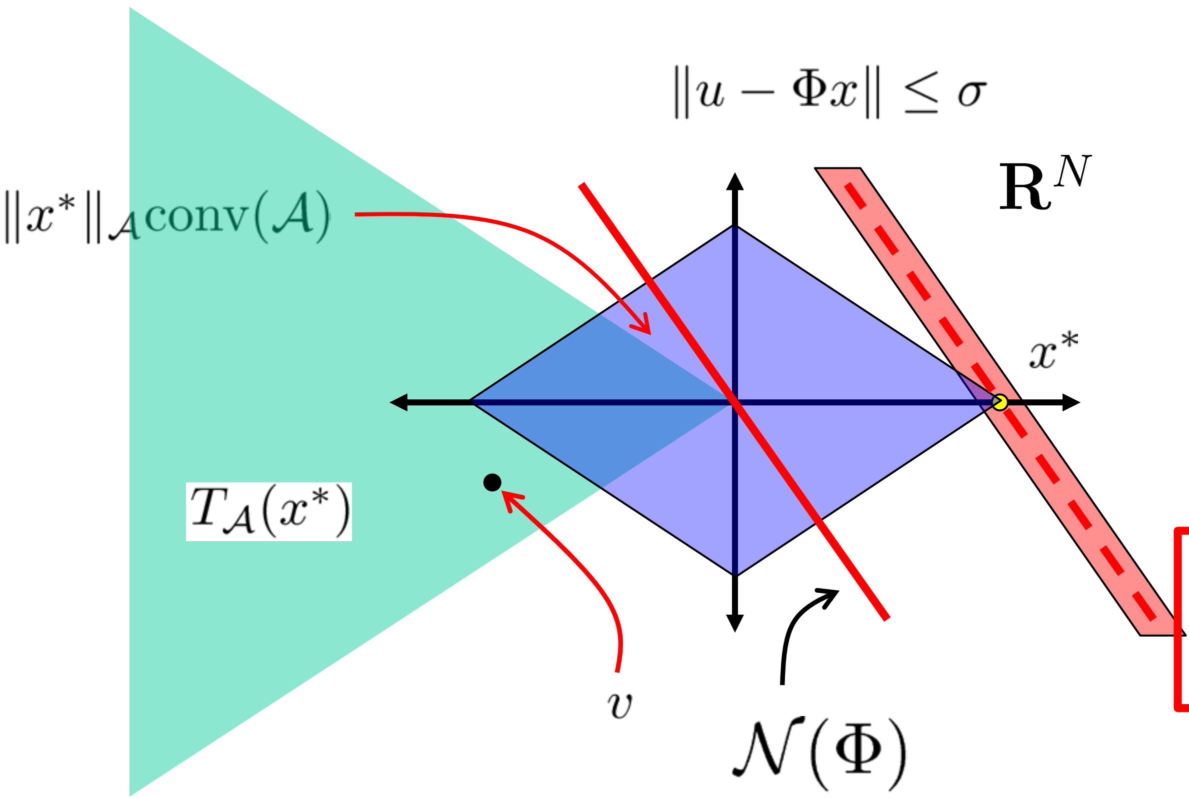
Key observation:

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\} \Rightarrow$$
$$x^* = \arg \min_{x:u=\Phi x} \|x\|_{\mathcal{A}}$$

Towards algorithms: a geometric perspective

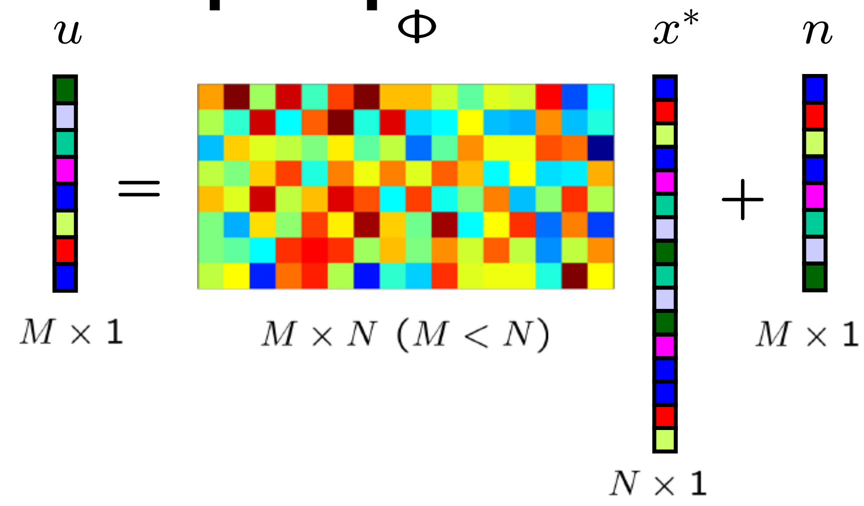
How about noise?



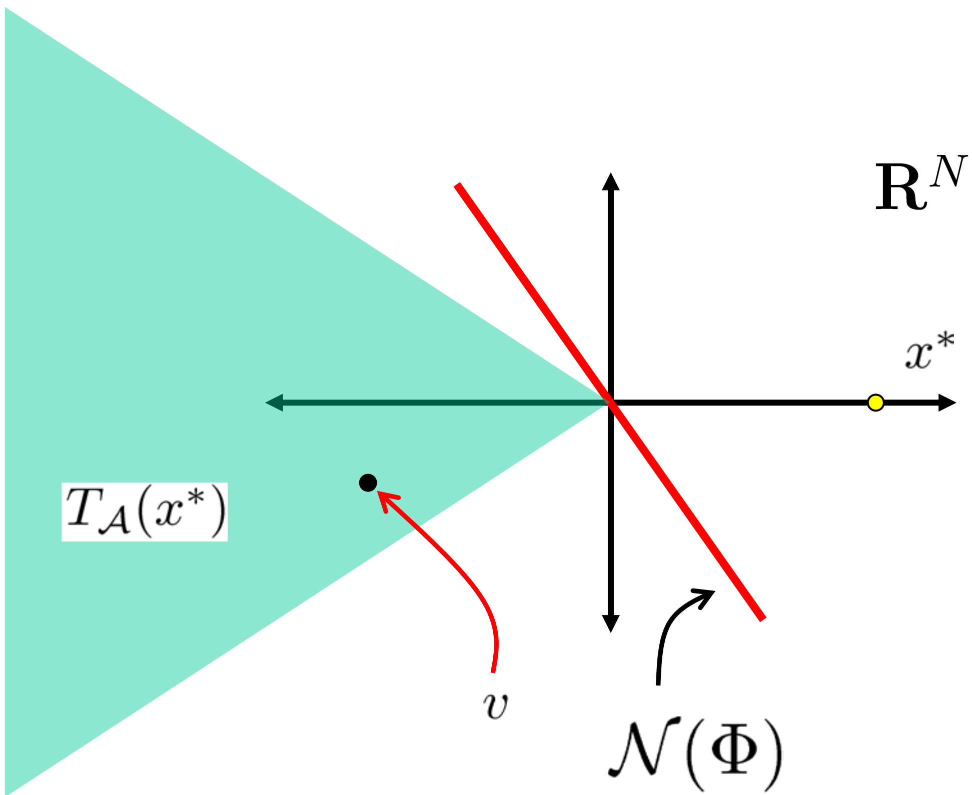


$$\widehat{x} = \arg \min_{x:\|u-\Phi x\| \leq \sigma} \|x\|_{\mathcal{A}}$$

Towards algorithms: a geometric perspective



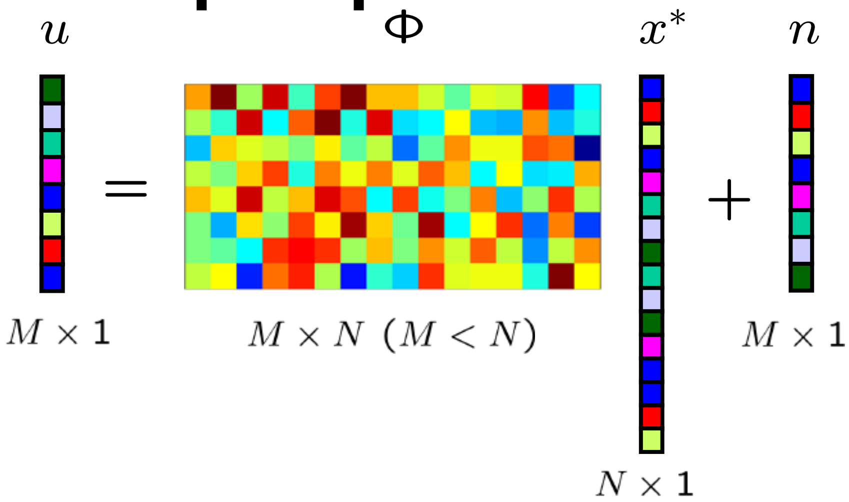
How about noise?



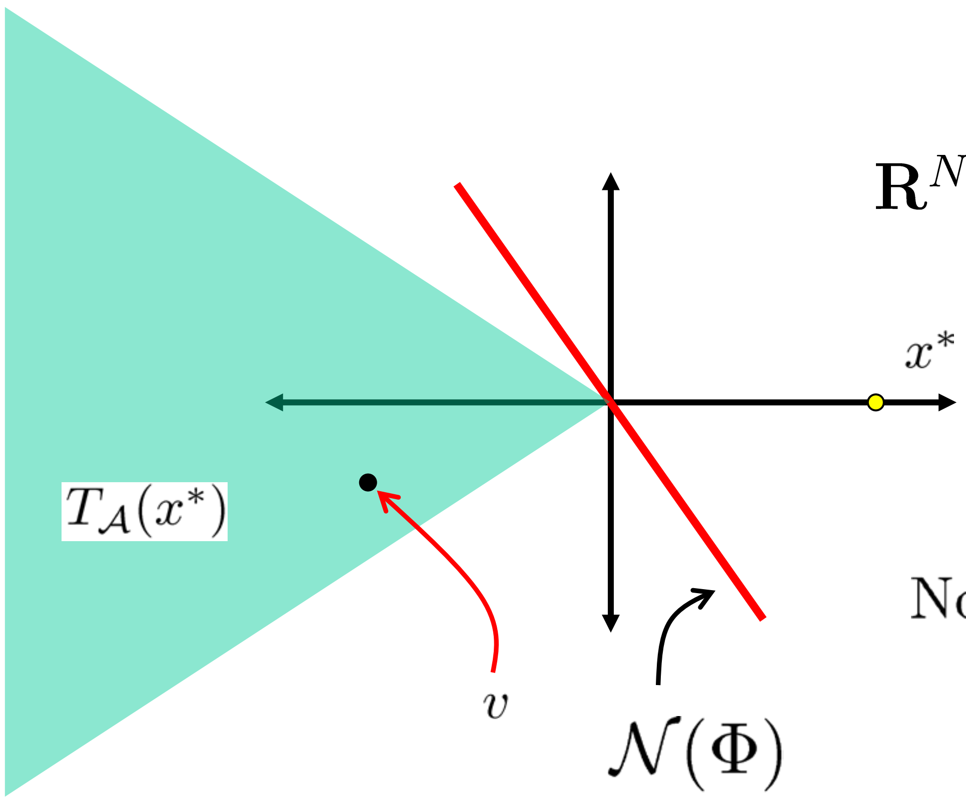
Stability assumption:

$$\|\Phi v\| \geq \epsilon \|v\|, \forall v \in T_{\mathcal{A}}(x^*)$$

Towards algorithms: a geometric perspective



How about noise?



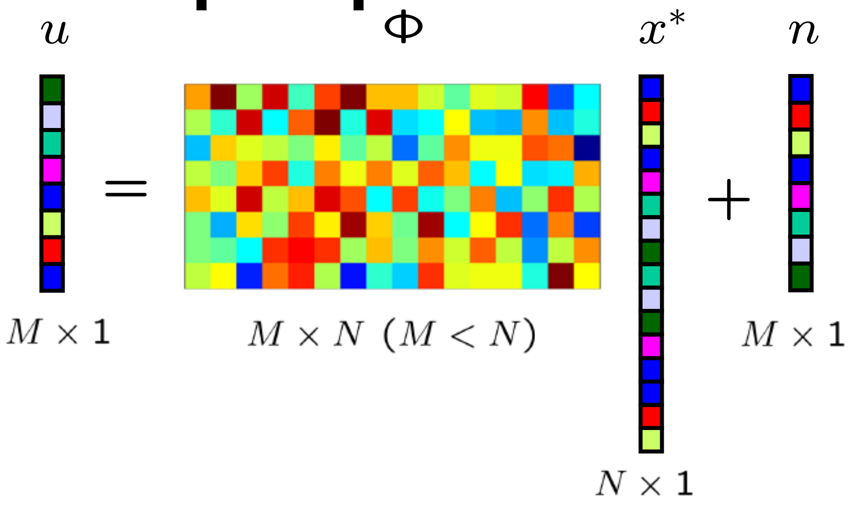
Stability assumption:

$$\|\Phi v\| \geq \epsilon \|v\|, \forall v \in T_{\mathcal{A}}(x^*)$$

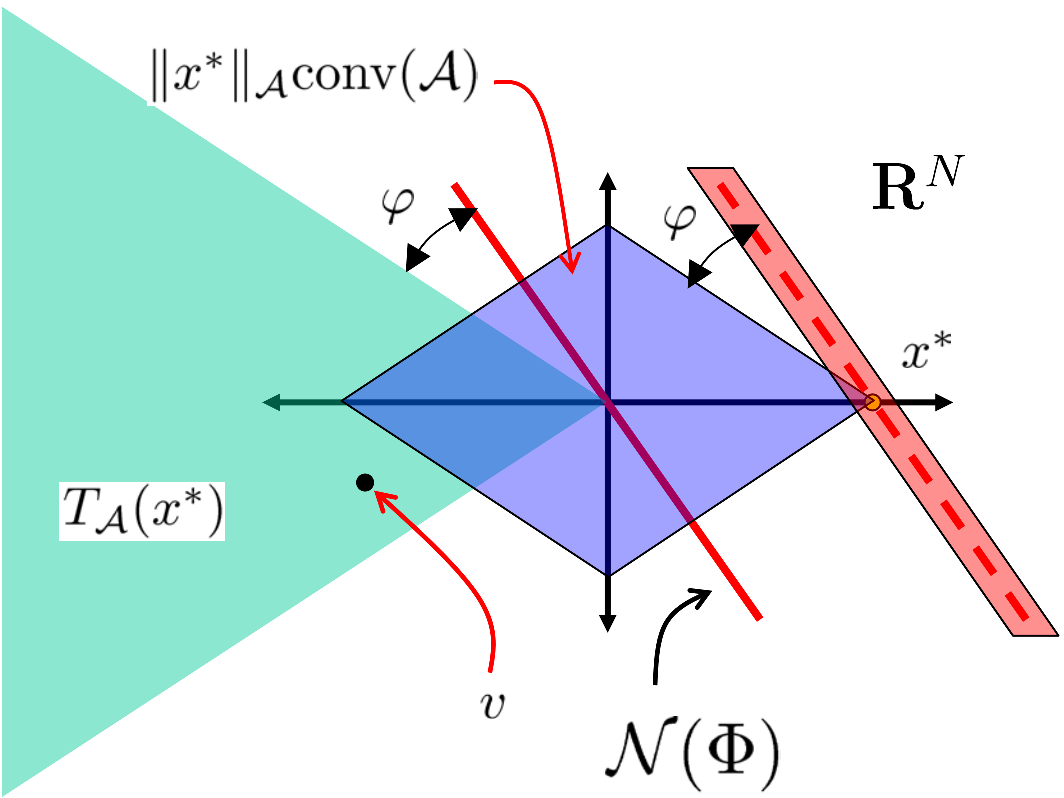
Note that if $\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\}$

$$\Rightarrow \|\Phi v\| > 0, \forall v \neq 0$$

Towards algorithms: a geometric perspective



How about noise?



Stability assumption:

$$\|\Phi v\| \geq \epsilon \|v\|, \forall v \in T_{\mathcal{A}}(x^*)$$

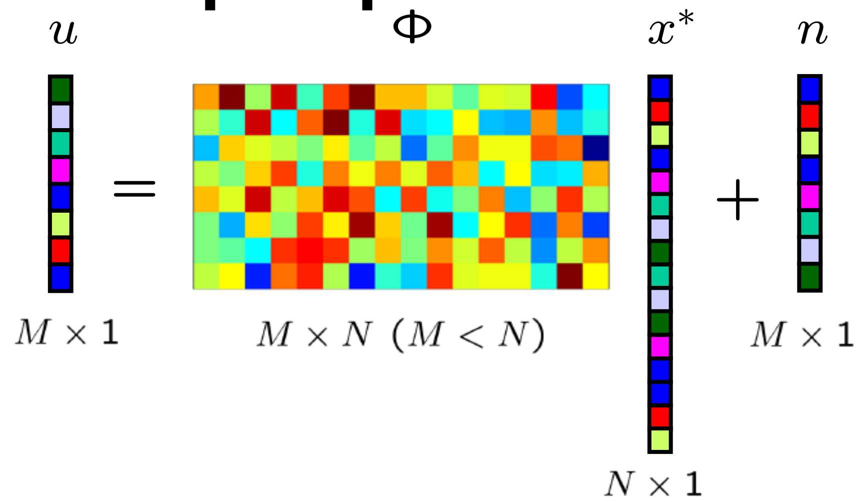
want epsilon large to minimize overlap between $\|x^*\|_{\mathcal{A}}\mathrm{conv}(\mathcal{A})$ and $\|u - \Phi x\| \leq \sigma$

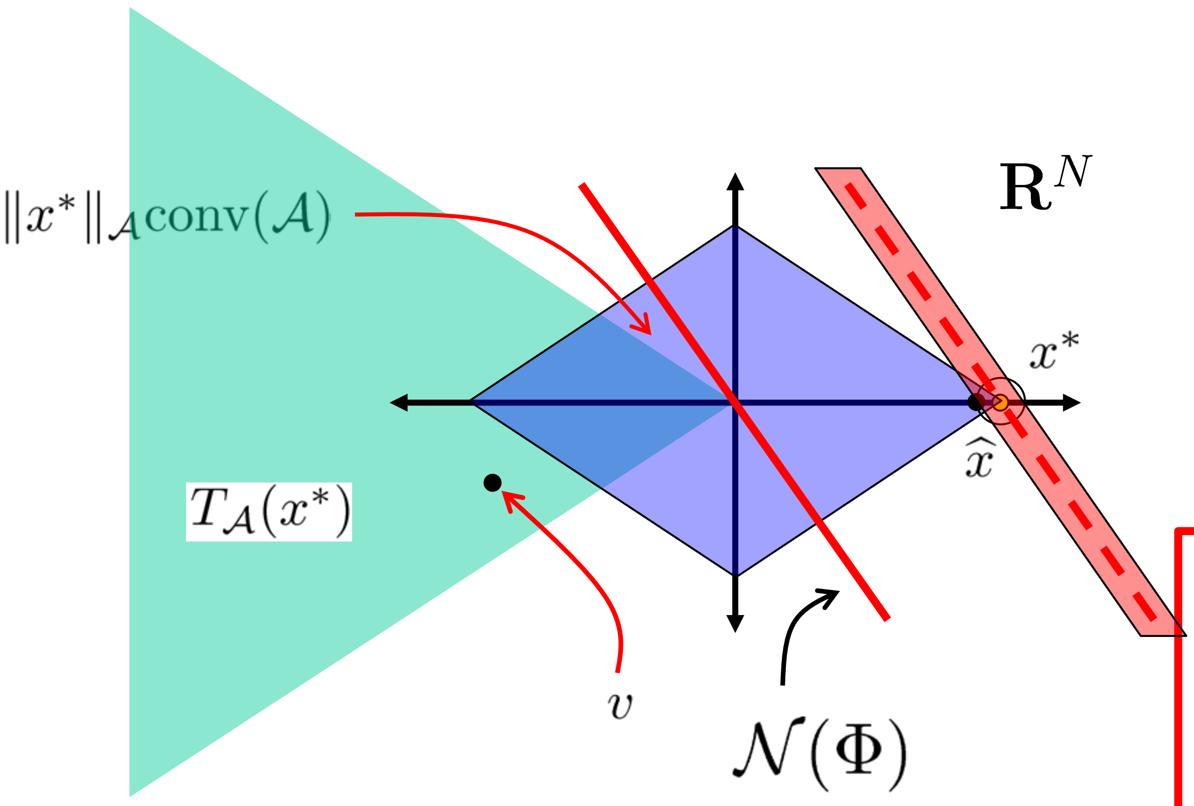
For this 2D example: $\|\Phi v\| \geq \|v\| \sin(\varphi) \min_i \|\Phi(i,:)\|$

Matlab notation

Towards algorithms: a geometric perspective

How about noise?





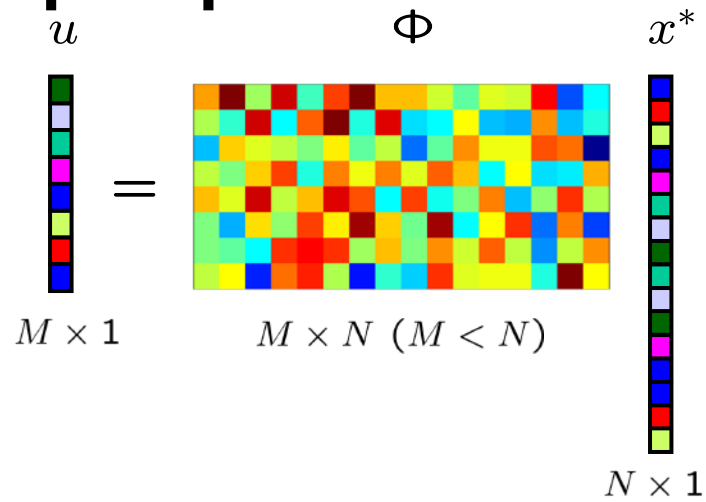
Stability assumption:

$$\|\Phi v\| \geq \epsilon \|v\|, \forall v \in T_{\mathcal{A}}(x^*)$$

$$\widehat{x} = \arg \min_{x: \|u - \Phi x\| \leq \sigma} \|x\|_{\mathcal{A}}$$

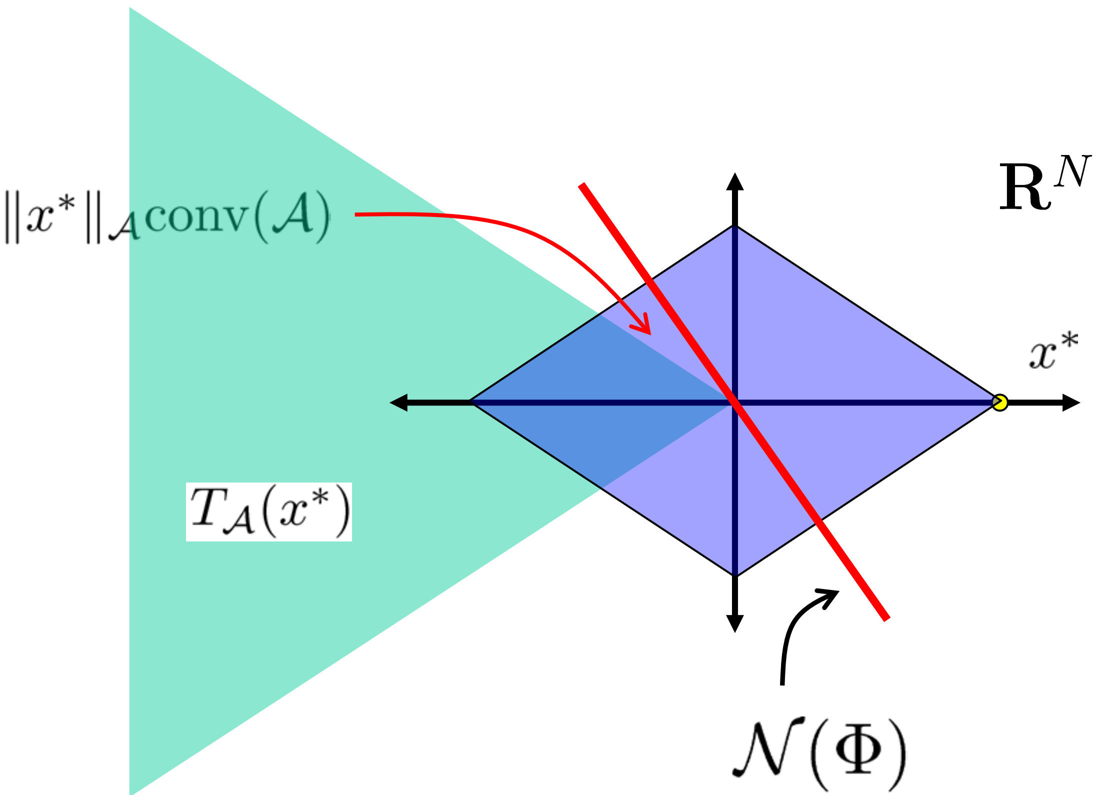
$$\Rightarrow \|x^* - \widehat{x}\| \leq \frac{2\sigma}{\epsilon}$$

Towards algorithms: a geometric perspective



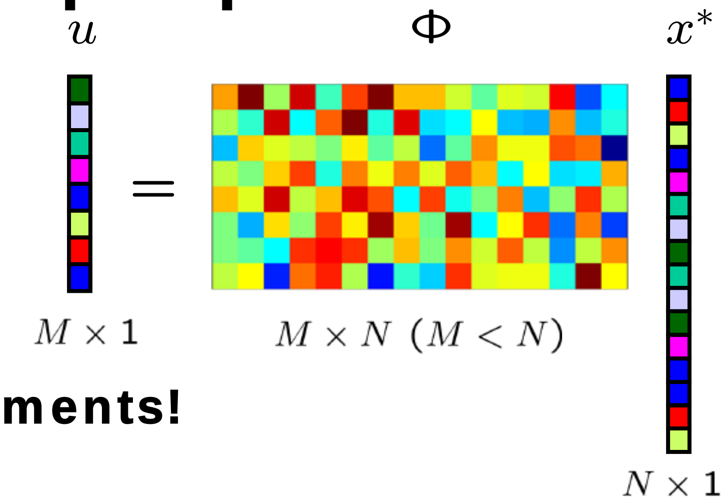
Can we guarantee the following?*

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\}$$



*without knowing x^*

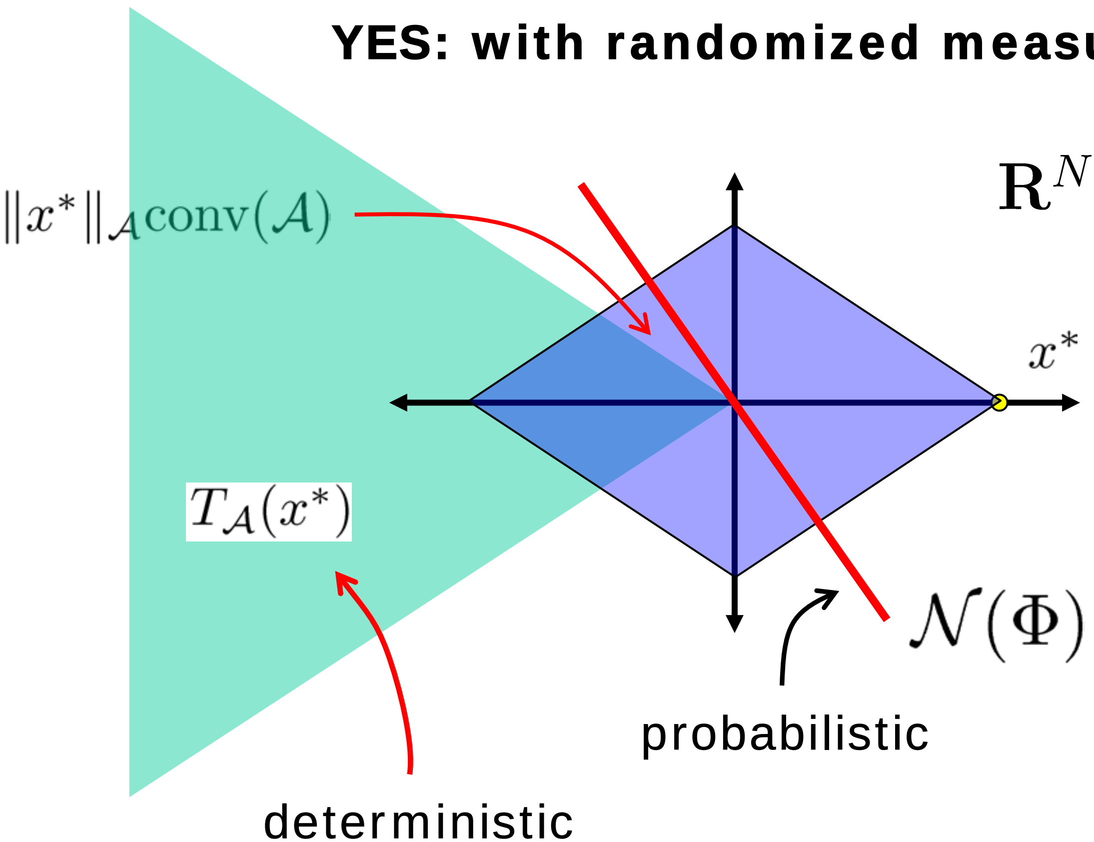
Towards algorithms: a geometric perspective



Can we guarantee the following?*

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\}$$

YES: with randomized measurements!



Gordon's Minimum Restricted Singular Values Theorem has a probabilistic characterization.

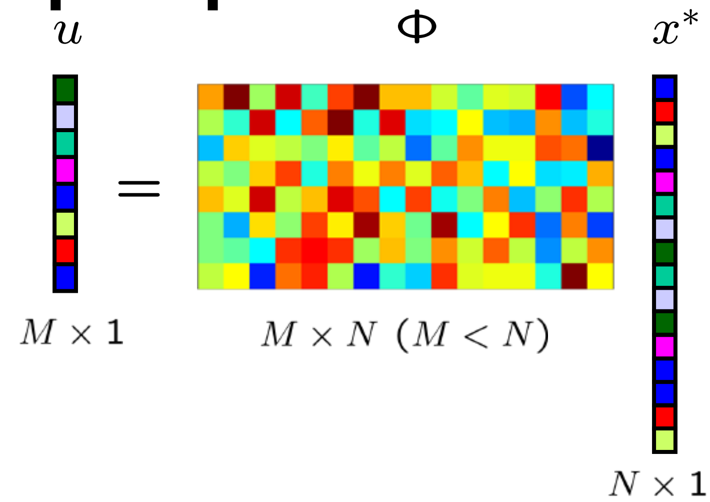
$$\mathrm{Prob}(\min_v \|\Phi v\| \geq \epsilon)$$

$$\forall v \in T_{\mathcal{A}}(x^*), \|v\| = 1$$

[Gordon 1988]

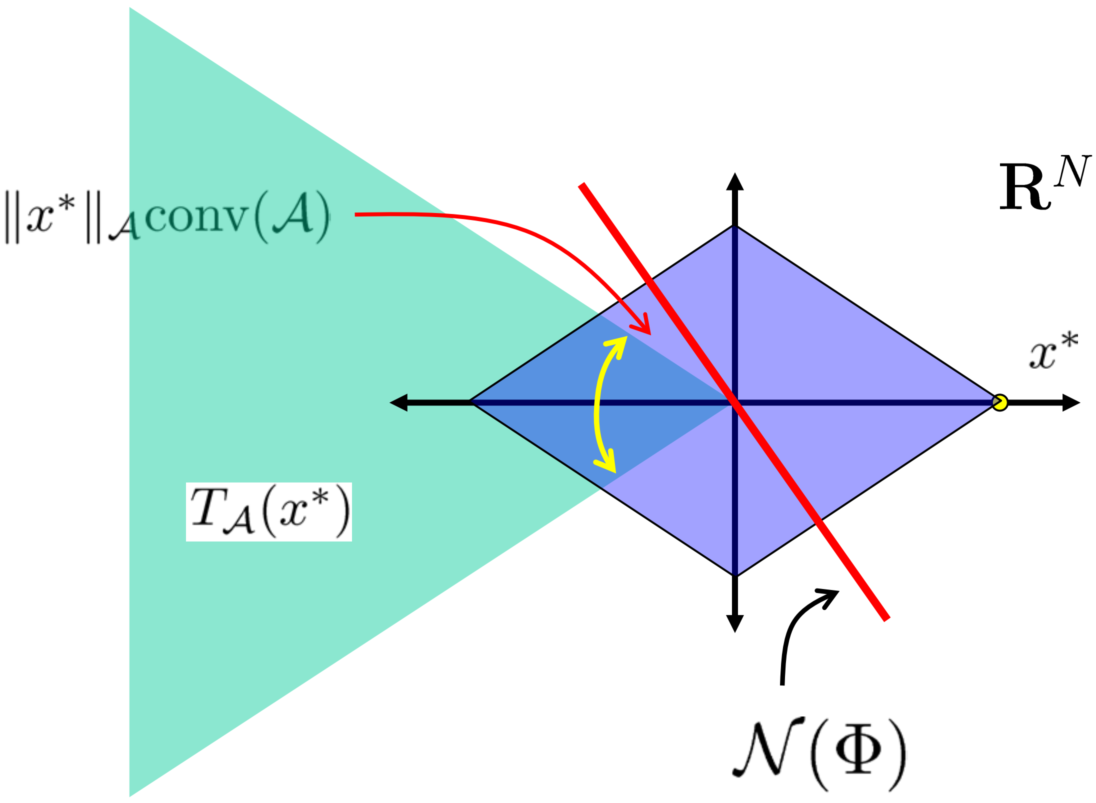
*without knowing x^*

Towards algorithms: a geometric perspective



Can we guarantee the following?*

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\}$$

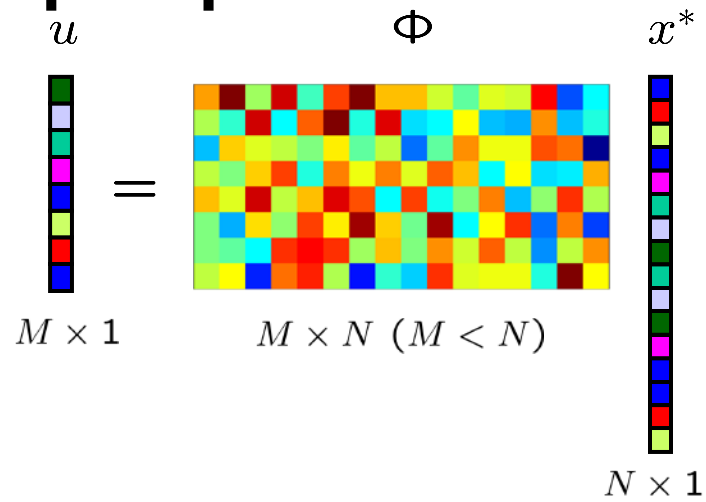


Gordon's Minimum Restricted Singular Values Theorem has a probabilistic characterization.

Key concept:
width of the tangent cone!

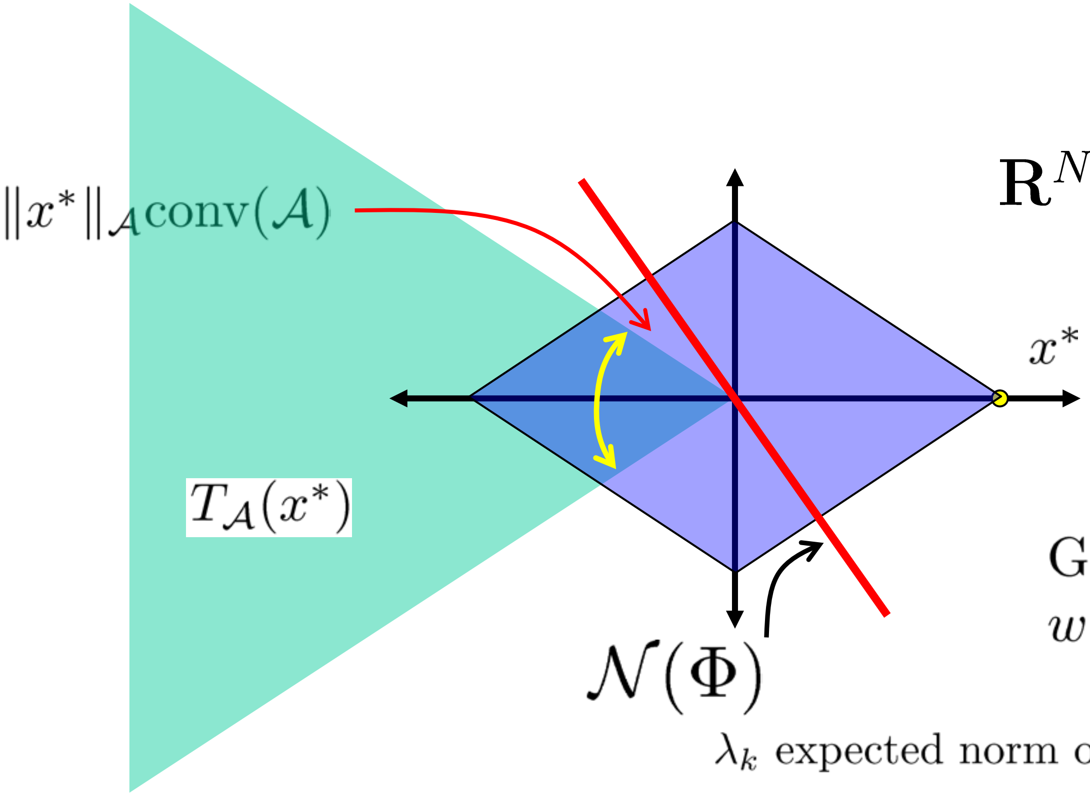
*without knowing x^*

Towards algorithms: a geometric perspective



Can we guarantee the following?*

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\}$$



Gordon's Minimum Restricted Singular Values Theorem has a probabilistic characterization.

Gaussian width of $S \subseteq \mathbb{R}^M$

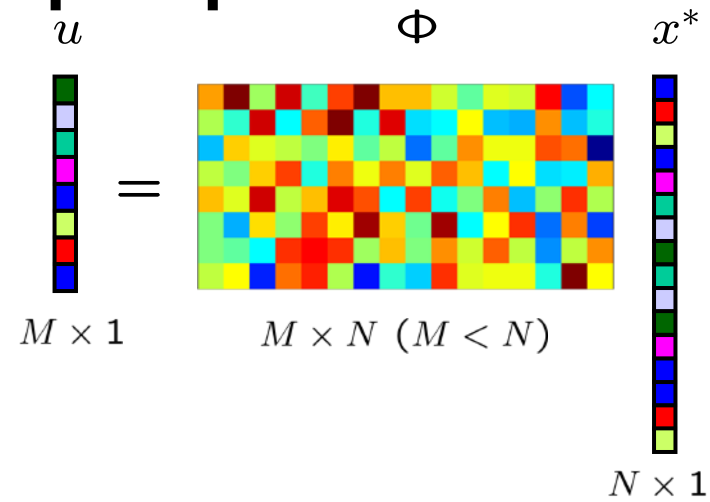
$$w(S) = \mathrm{E}\left[\sup_{z \in S} g^T z\right]; \; g \sim \mathcal{N}(0, I)$$

λ_k expected norm of a k-dimensional Gaussian random vector:

$$\lambda_k = \sqrt{\mathrm{E}\left[\sum_{i=1}^{k} g_i^2\right]} = \frac{\sqrt{2}\Gamma((k+1)/2)}{k/2}$$

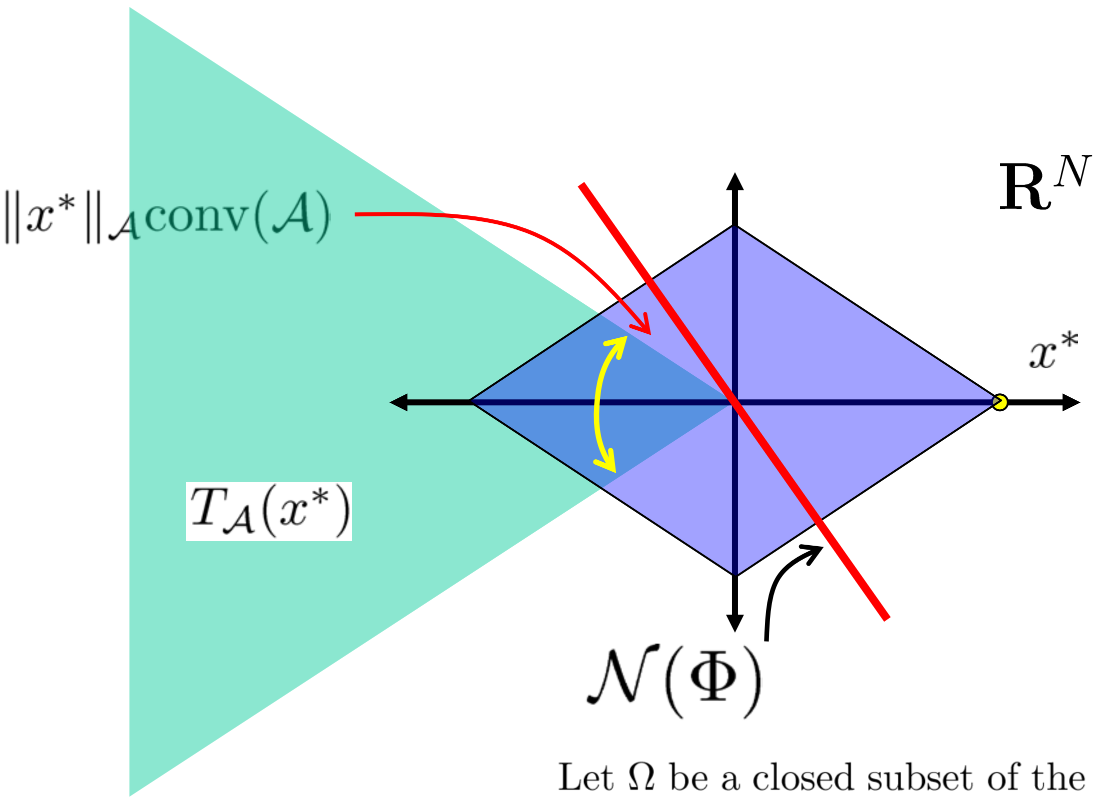
*without knowing x^*

Towards algorithms: a geometric perspective



Can we guarantee the following?*

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\}$$



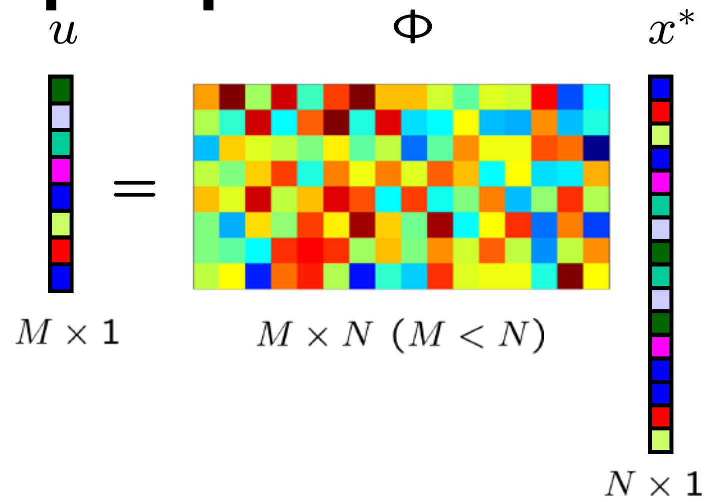
Gordon's Minimum Restricted Singular Values Theorem has a probabilistic characterization.

Let Ω be a closed subset of the unit sphere and A be an $M \times N$ matrix with iid $\mathcal{N}(0,1)$ entries. Then, if $\lambda_k \geq w(\Omega) + \epsilon$:

$$P\left[\min_{z \in \Omega} \|Az\|_2 \geq \epsilon\right] \geq 1 - \frac{1}{2} \mathrm{e}^{-\frac{1}{18}(\lambda_k - w(\Omega)\epsilon)^2}$$

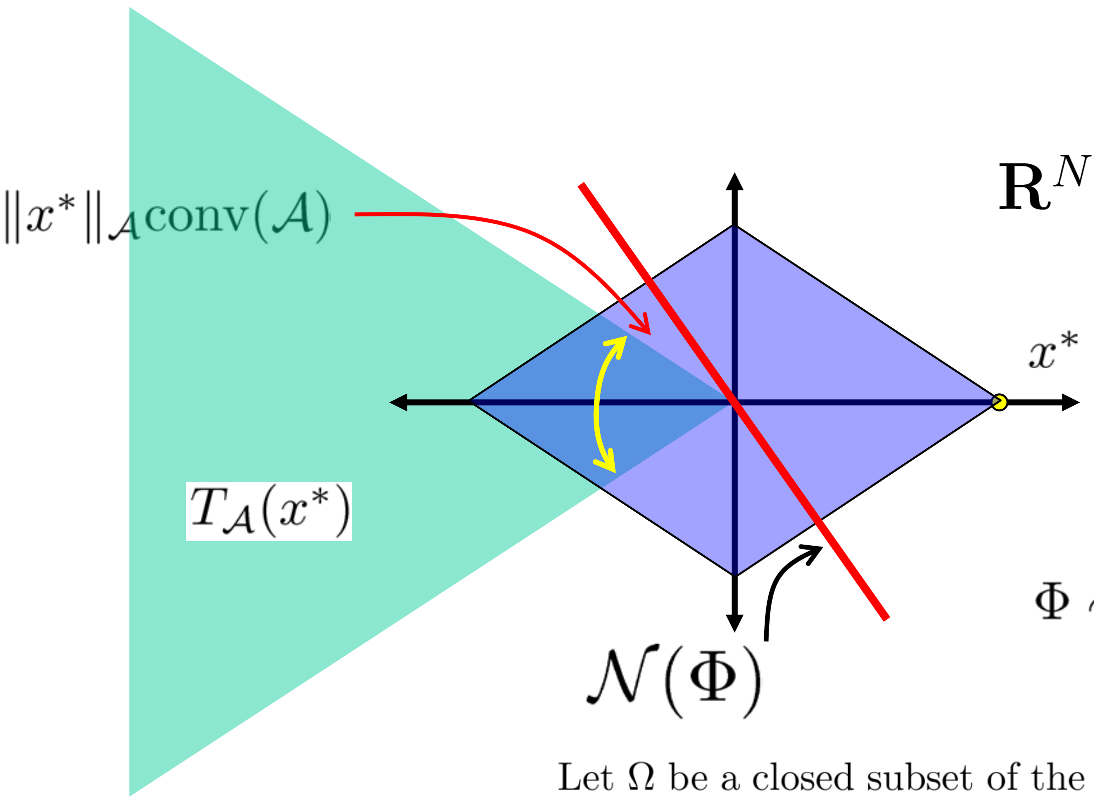
*without knowing x^*

Towards algorithms: a geometric perspective



Can we guarantee the following?*

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\}$$



Gordon's Minimum Restricted Singular Values Theorem has a probabilistic characterization.

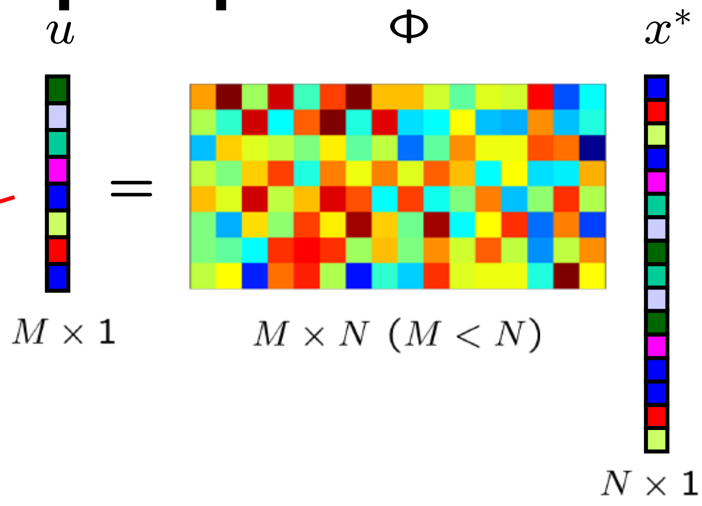
$$\Phi \sim_{\text{iid}} \mathcal{N}(0, 1/M), \Omega = T_{\mathcal{A}}(x^*) \cap \mathbb{S}^{N-1}$$

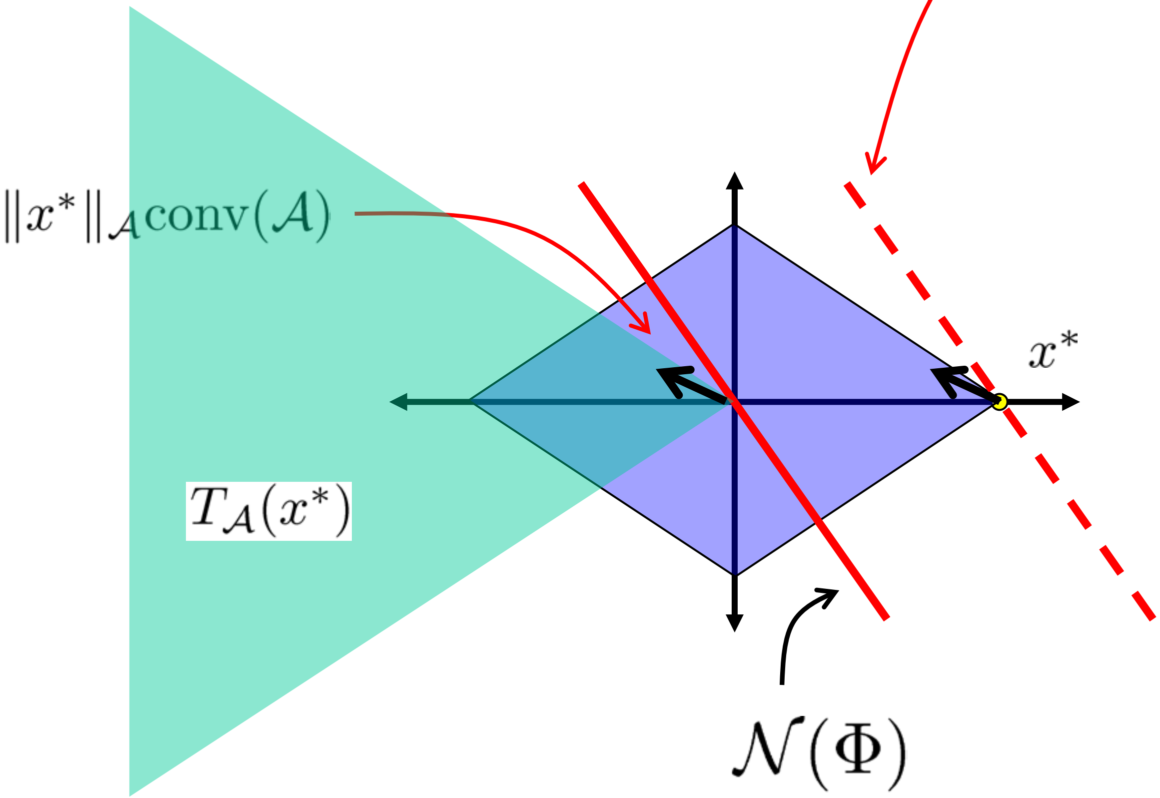
Let Ω be a closed subset of the unit sphere and A be an $M \times N$ matrix with iid $\mathcal{N}(0,1)$ entries. Then, if $\lambda_k \geq w(\Omega) + \epsilon$:

$$P\left[\min_{z\in\Omega} \|Az\|_2 \geq \epsilon\right] \geq 1 - \frac{1}{2}\mathrm{e}^{-\frac{1}{18}(\lambda_k - w(\Omega)-\epsilon)^2}$$

*without knowing x^*

Towards algorithms: a geometric perspective





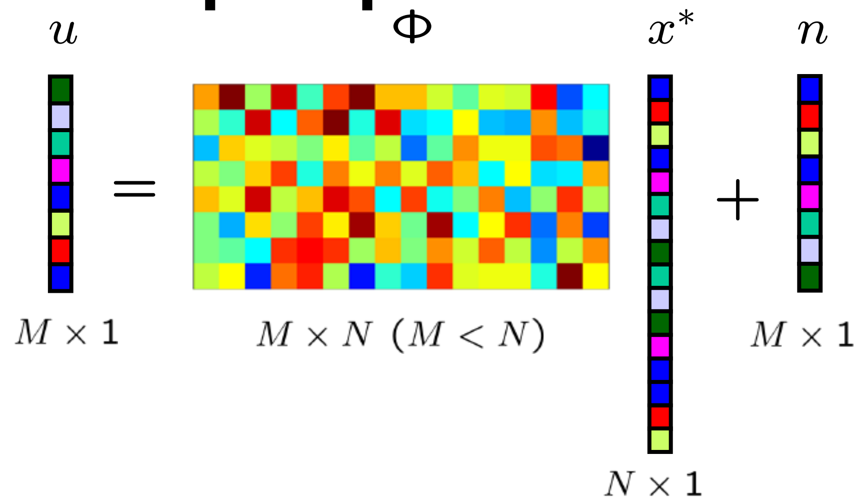
Key observation:

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\} \Rightarrow$$

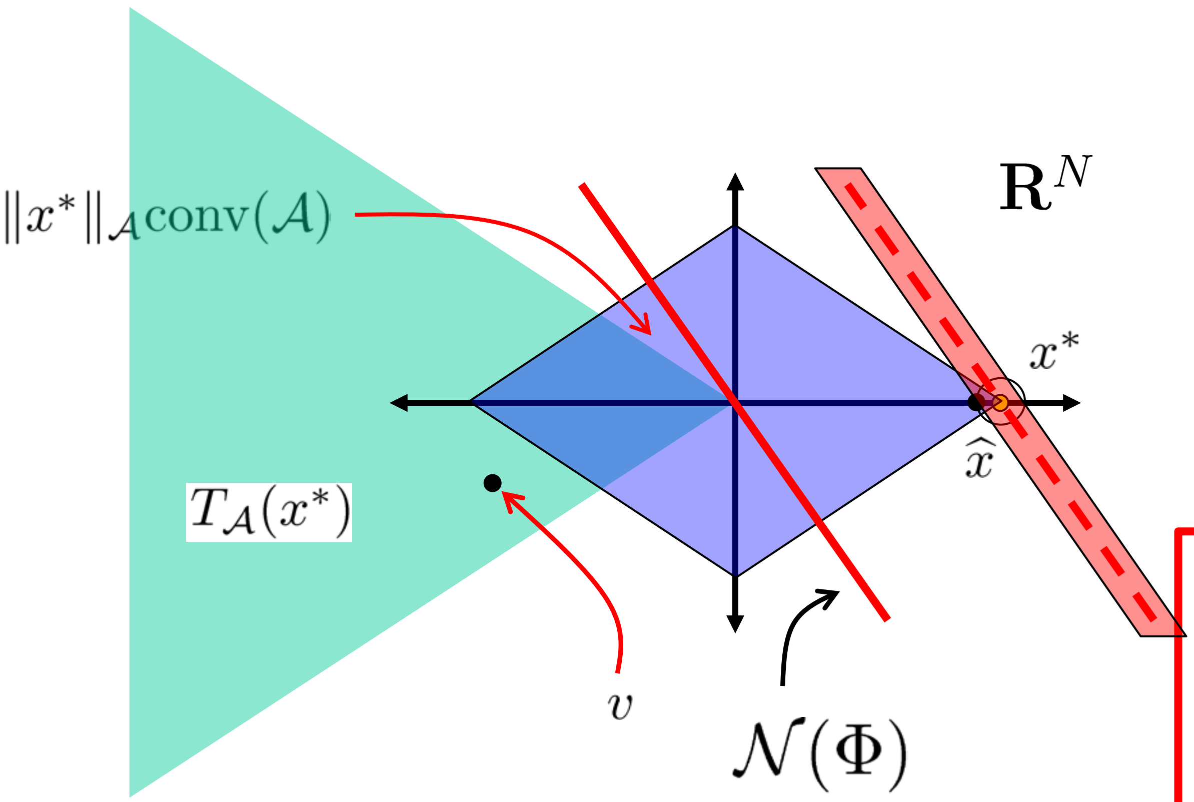
$$x^* = \arg \min_{x:u=\Phi x} \|x\|_{\mathcal{A}}$$

$$M \geq w(\Omega)^2 + \mathcal{O}(1)$$

Towards algorithms: a geometric perspective



How about noise?



Stability assumption:

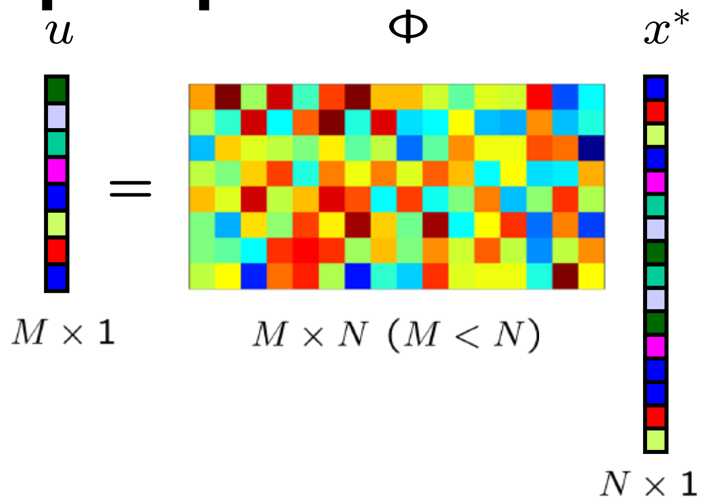
$$\|\Phi v\| \geq \epsilon \|v\|, \forall v \in T_{\mathcal{A}}(x^*)$$

$$\widehat{x} = \arg \min_{x:\|u - \Phi x\| \leq \sigma} \|x\|_{\mathcal{A}}$$

$$\Rightarrow \|x^* - \widehat{x}\| \leq \frac{2\sigma}{\epsilon}$$

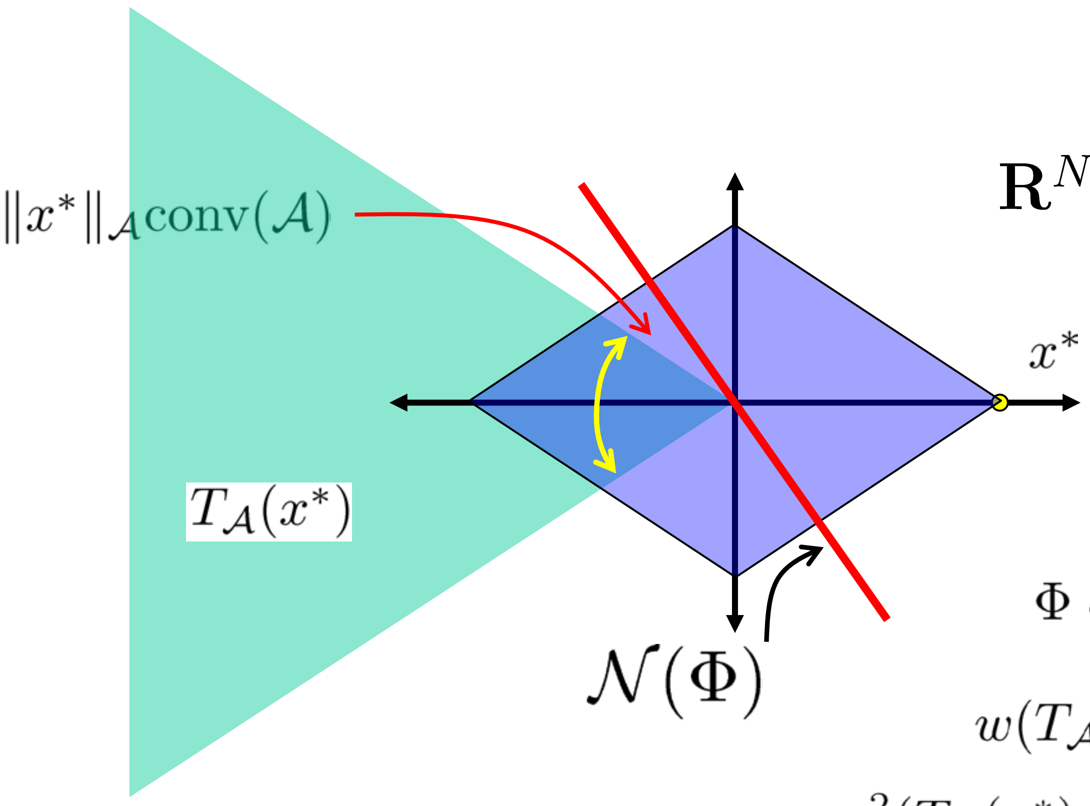
$$M \geq \frac{w(\Omega)^2}{(1-\epsilon)^2} + \mathcal{O}(1)$$

Towards algorithms: a geometric perspective



Can we guarantee the following?*

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\}$$



Gordon's Minimum Restricted Singular Values Theorem has a probabilistic characterization.

$$g \sim_{\text{iid}} \mathcal{N}(0,1)$$

$$\Phi \sim_{\text{iid}} \mathcal{N}(0, 1/M), \Omega = T_{\mathcal{A}}(x^*) \cap \mathbb{S}^{N-1}$$

$$w(T_{\mathcal{A}}(x^*) \cap \mathbb{S}^{N-1}) \leq \mathrm{E}_g \left[\mathrm{dist}\left(g, T^{\circ}_{\mathcal{A}}(x^*)\right)\right]$$

$$w^2(T_{\mathcal{A}}(x^*) \cap \mathbb{S}^{N-1}) + w^2(T^{\circ}_{\mathcal{A}}(x^*) \cap \mathbb{S}^{N-1}) \leq N$$

$$w(T_{\mathcal{A}}(x^*) \cap \mathbb{S}^{N-1}) \leq \sqrt{\log\left(\frac{4}{\mathrm{vol}(T^{\circ}_{\mathcal{A}}(x^*) \cap \mathbb{S}^{N-1})}\right)}$$

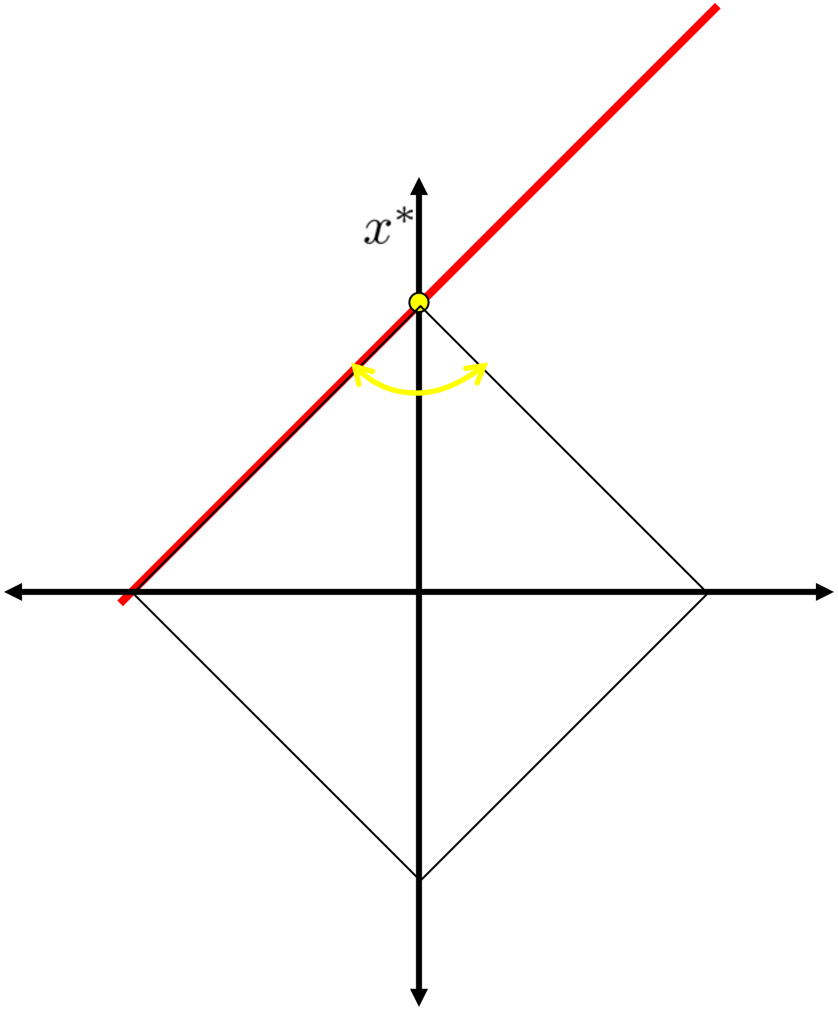
$N \geq 9$

*without knowing x^*

Towards algorithms: a geometric perspective

Can we guarantee the following?*

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\}$$



$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$$

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\} \text{ w.p. } 1/2$$

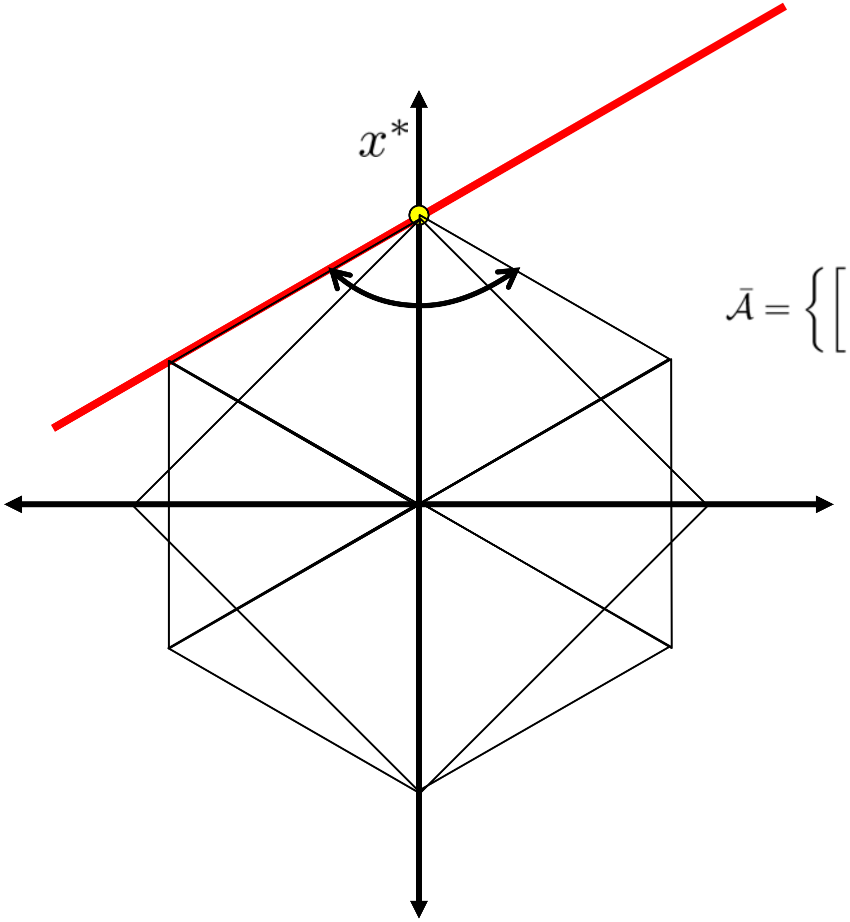
$$\Rightarrow x^* = \arg \min_{x : u = \Phi x} \|x\|_1$$

*without knowing 1-sparse x^* and 1-random measurement

Towards algorithms: a geometric perspective

Can we guarantee the following?*

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\}$$



$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$$

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\} \text{ w.p. } 1/2$$

$$\Rightarrow x^* = \arg \min_{x:u=\Phi x} \|x\|_1$$

$$\bar{\mathcal{A}} = \left\{ \begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -\sqrt{3}/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} -\sqrt{3}/2 \\ -1/2 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \begin{bmatrix} \sqrt{3}/2 \\ -1/2 \end{bmatrix} \right\}$$

$$\mathcal{N}(\Phi) \cap T_{\bar{\mathcal{A}}}(x^*) = \{0\} \text{ w.p. } 1/3$$

$$\Rightarrow x^* = \arg \min_{x:u=\Phi x} \|x\|_{\bar{\mathcal{A}}}$$

*without knowing 1-sparse x^* and 1-random measurement

Towards algorithms: a geometric perspective

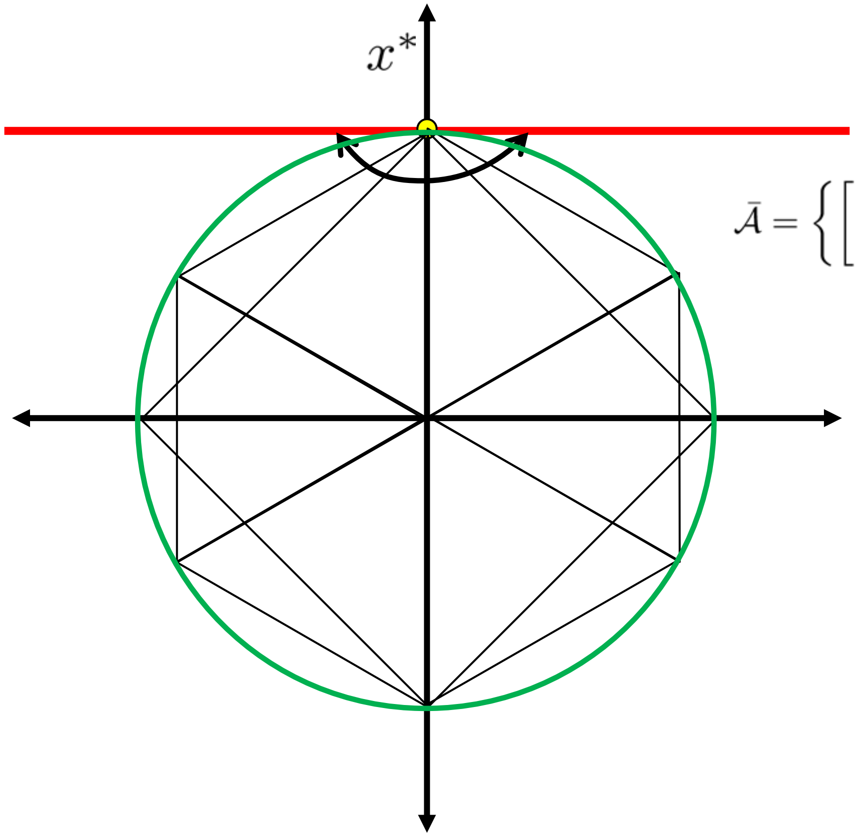
Can we guarantee the following?*

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\}$$

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$$

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\} \text{ w.p. } 1/2$$

$$\Rightarrow x^* = \arg \min_{x:u=\Phi x} \|x\|_1$$



$$\bar{\mathcal{A}} = \left\{ \begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -\sqrt{3}/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} -\sqrt{3}/2 \\ -1/2 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \begin{bmatrix} \sqrt{3}/2 \\ -1/2 \end{bmatrix} \right\}$$

$$\mathcal{N}(\Phi) \cap T_{\bar{\mathcal{A}}}(x^*) = \{0\} \text{ w.p. } 1/3$$

$$\Rightarrow x^* = \arg \min_{x:u=\Phi x} \|x\|_{\bar{\mathcal{A}}}$$

$$\tilde{\mathcal{A}} = \{\|x\|_2 = 1\}$$

$$\mathcal{N}(\Phi) \cap T_{\tilde{\mathcal{A}}}(x^*) = \{0\} \text{ w.p. } 0$$

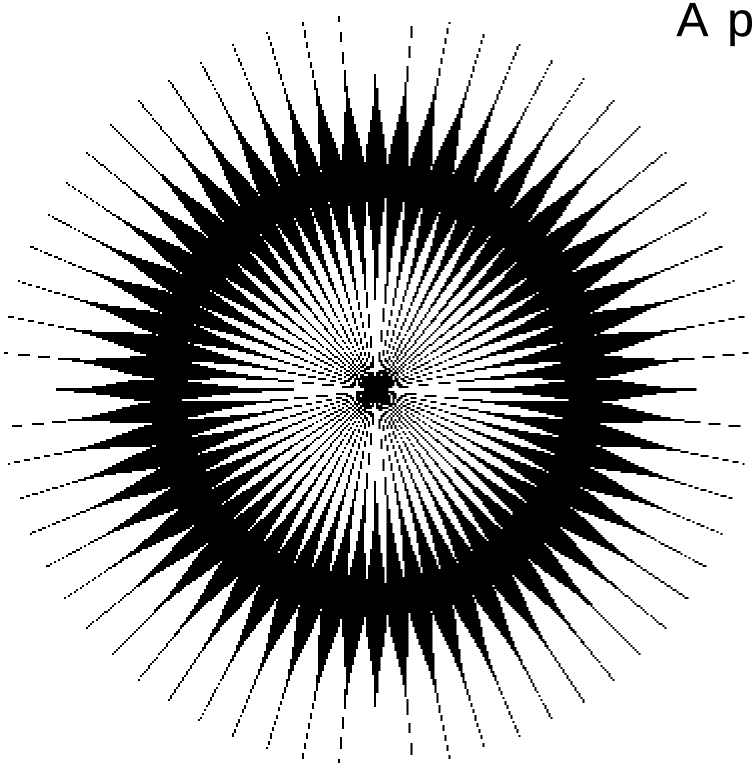
$$\Rightarrow x^* = \arg \min_{x:u=\Phi x} \|x\|_2$$

*without knowing 1-sparse x^* and 1-random measurement

Towards algorithms: a geometric perspective

Can we guarantee the following?*

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\}$$



A projected 6D hypercube with 64 vertices

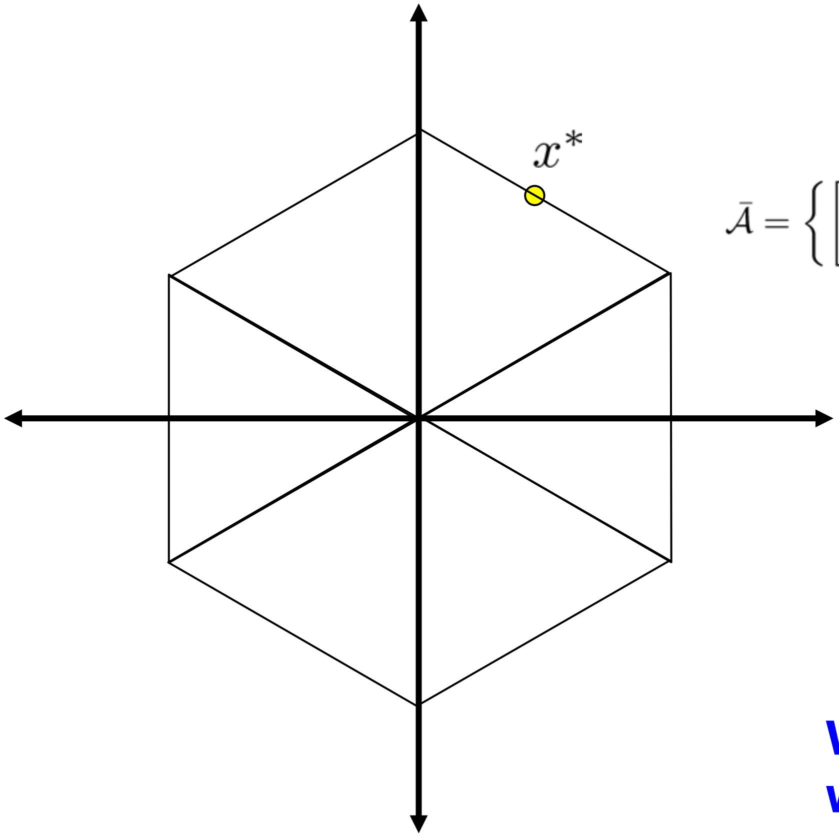
Blessing-of-dimensionality!

http://www.agrell.info/erik/chalmers/hypercubes/

Towards algorithms: a geometric perspective

Pop-quiz:

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\}$$



$$\bar{\mathcal{A}} = \left\{ \begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -\sqrt{3}/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} -\sqrt{3}/2 \\ -1/2 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \begin{bmatrix} \sqrt{3}/2 \\ -1/2 \end{bmatrix} \right\}$$

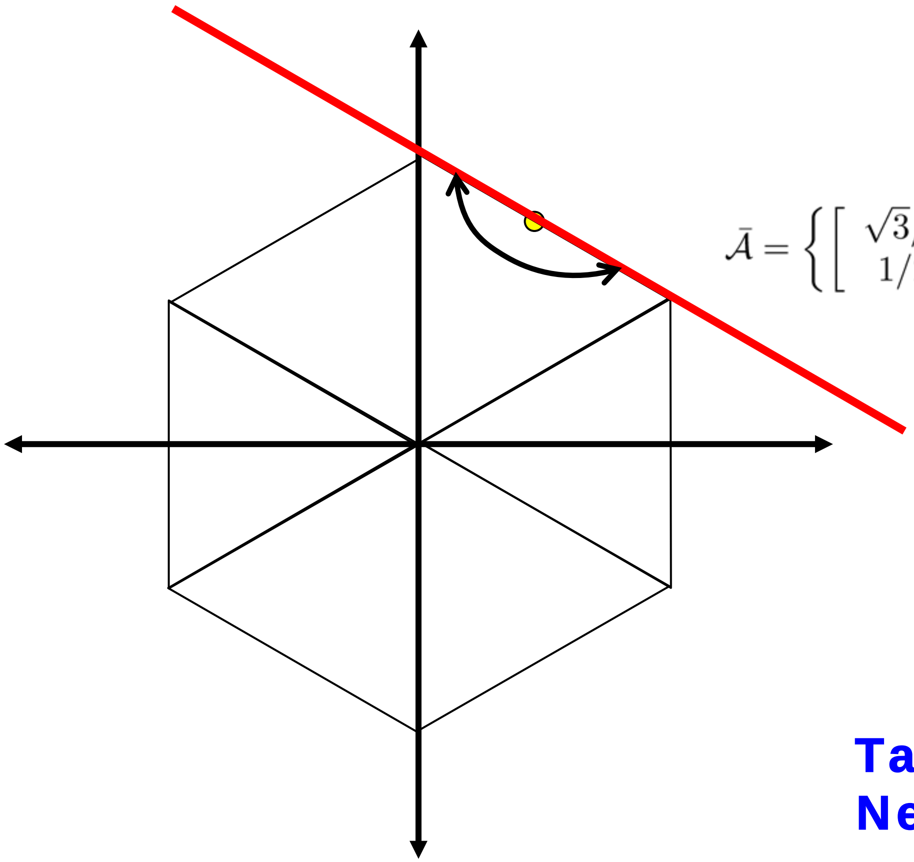
$$\mathcal{N}(\Phi) \cap T_{\bar{\mathcal{A}}}(x^*) = \{0\} \text{ w.p. ???}$$
$$\Rightarrow x^* = \arg \min_{x:u=\Phi x} \|x\|_{\bar{\mathcal{A}}}$$

What is the probability that we can determine a 2-sparse x* with 1-random measurement?

Towards algorithms: a geometric perspective

Pop-answer:

$$\mathcal{N}(\Phi) \cap T_{\mathcal{A}}(x^*) = \{0\}$$



$$\bar{\mathcal{A}} = \left\{ \begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -\sqrt{3}/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} -\sqrt{3}/2 \\ -1/2 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \begin{bmatrix} \sqrt{3}/2 \\ -1/2 \end{bmatrix} \right\}$$

$$\mathcal{N}(\Phi) \cap T_{\bar{\mathcal{A}}}(x^*) = \{0\} \text{ w.p. } 0$$

$$\Rightarrow x^* = \arg \min_{x : u = \Phi x} \|x\|_{\bar{\mathcal{A}}}$$

Tangent cone is too wide!
Need at least 2 measurements!

Take home messages

Underlying Model	Atomic Norm	Gaussian Measurements
K-sparse vector in $\mathbb{R}^N$	ℓ_1-norm	$(2K+1)\log(N-K)$
$N \times N$ rank-R matrix	nuclear norm	$3R(2N-R)+2(N-R-R^2)$
sign vector $\{\pm 1\}^N$	ℓ_∞-norm	$N/2$
$N \times N$-perm. matrix	Birkoff polytope norm	$9N\log(N)$
$N \times N$ orth. matrix	spectral norm	$(3N^2-N)/4$

[Chandrasekaran et al. 2010]

convex polytope <> atomic norm

– geometry (and algebra) of representations in **high dimensions**

geometric perspective <> convex criteria

– convex optimization algorithms in **high dimensions**

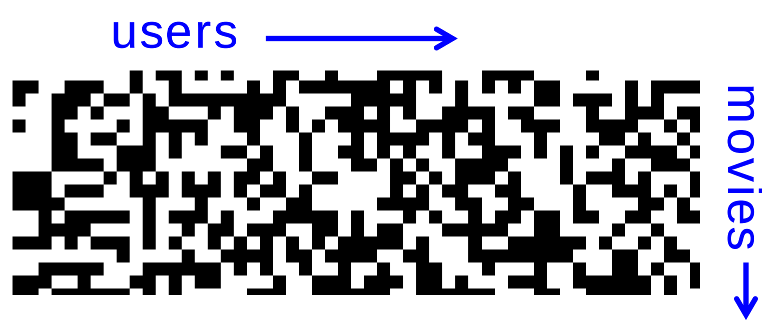
tangent cone width <> # of randomized samples

– probabilistic concentration-of-measures in **high dimensions**

Back to the initial example

- Matrix completion for Netflix   17770 movies x 480189 users

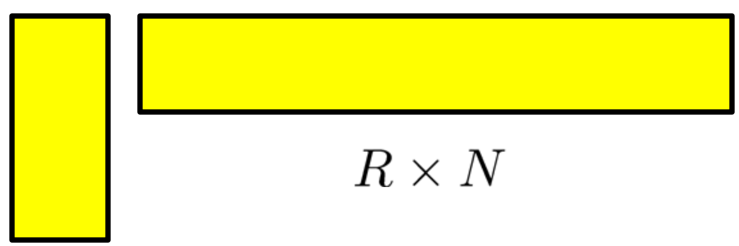
$X =$



$M \times N$

- What is low-rank?

$X =$



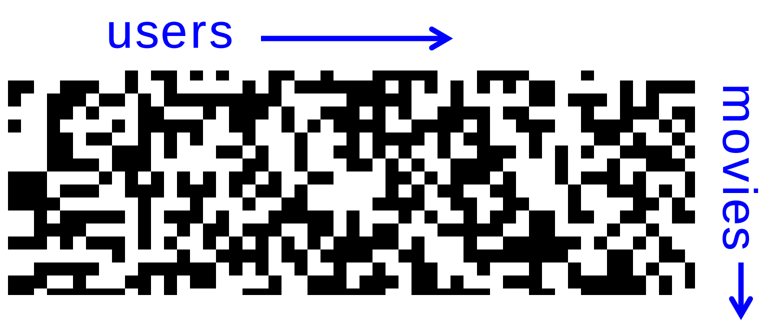
$R \times N$

$M \times R$

$$R \ll \min\{M, N\}$$

Back to the initial example

- Matrix completion for Netflix   17770 movies x 480189 users

$X =$ 

$$M \times N$$

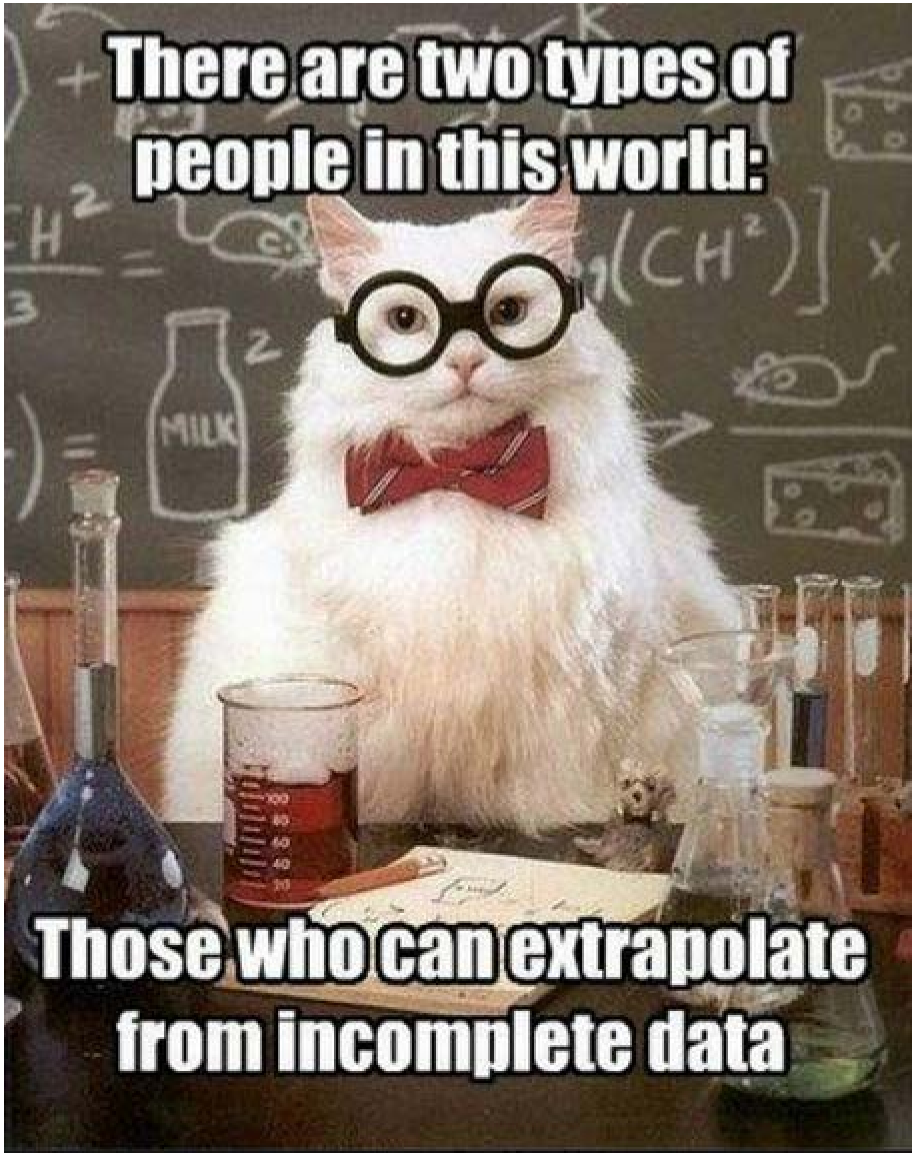
- What does the simple low-rank assumption buy?

Leaderboard   Display top 20 leaders.

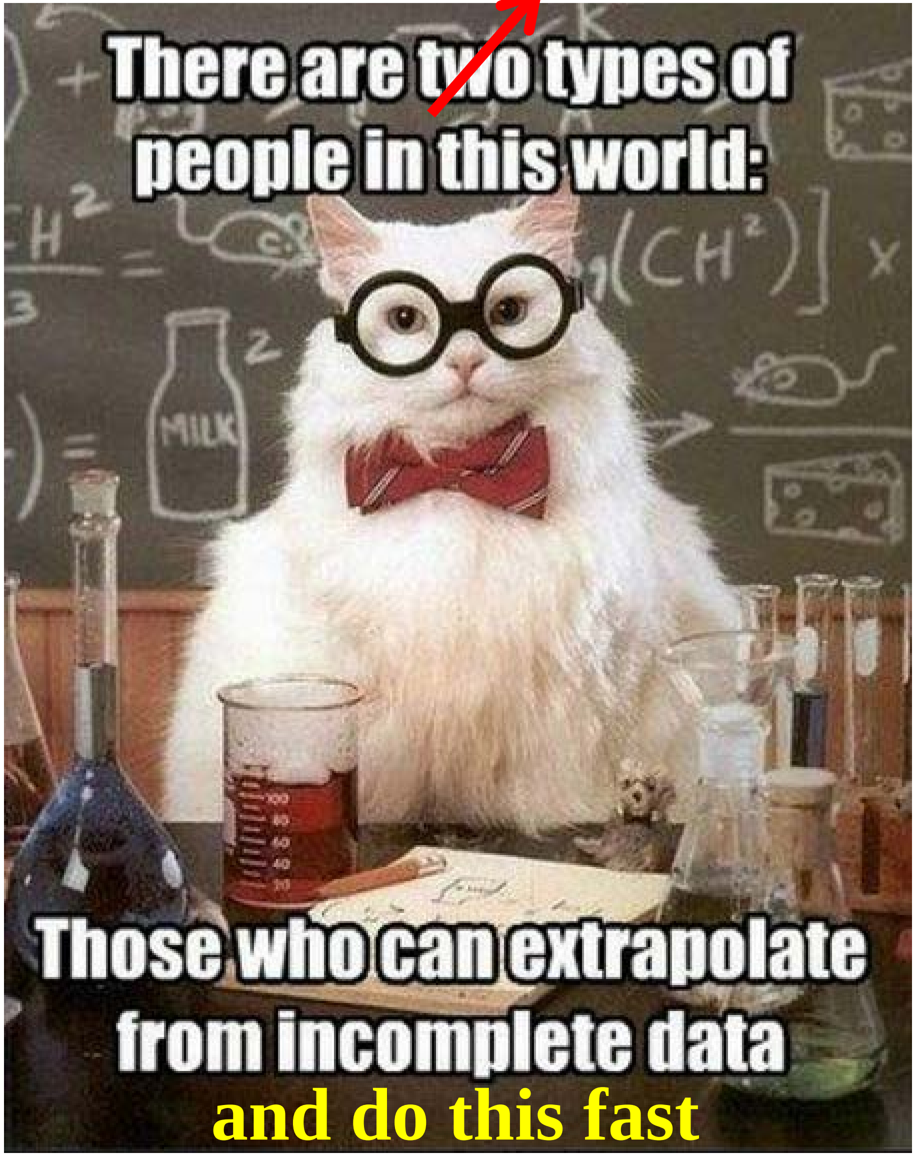
Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	The Ensemble	0.8553	10.10	2009-07-26 18:38:22
2	BellKor's Pragmatic Chaos	0.8554	10.09	2009-07-26 18:18:28
Grand Prize - RMSE <= 0.8563				
3	Grand Prize Team	0.8571	9.91	2009-07-24 13:07:49
4	Opera Solutions and Vandelay United	0.8573	9.89	2009-07-25 20:05:52
5	Vandelay Industries !	0.8579	9.83	2009-07-26 02:49:53
6	PragmaticTheory	0.8582	9.80	2009-07-12 15:09:53
7	BellKor in BigChaos	0.8590	9.71	2009-07-26 12:57:25
8	Dace_	0.8603	9.58	2009-07-24 17:18:43
9	Opera Solutions	0.8611	9.49	2009-07-26 18:02:08
10	BellKor	0.8612	9.48	2009-07-26 17:19:11
11	BigChaos	0.8613	9.47	2009-06-23 23:06:52
12	Feeds2	0.8613	9.47	2009-07-24 20:06:46

⋮	⋮	⋮	⋮	⋮
50	amgl	0.8897	6.49	2007-12-23 18:44:03
51	Remco	0.8899	6.46	2007-04-04 06:16:56
52	mdg	0.8900	6.45	2007-12-23 18:54:46
53	JustWithSVD	0.8900	6.45	2008-02-14 16:17:54

quite a lot of extrapolation power!


There are two types of
people in this world:
Those who can extrapolate
from incomplete data

three



with theoretical guarantees

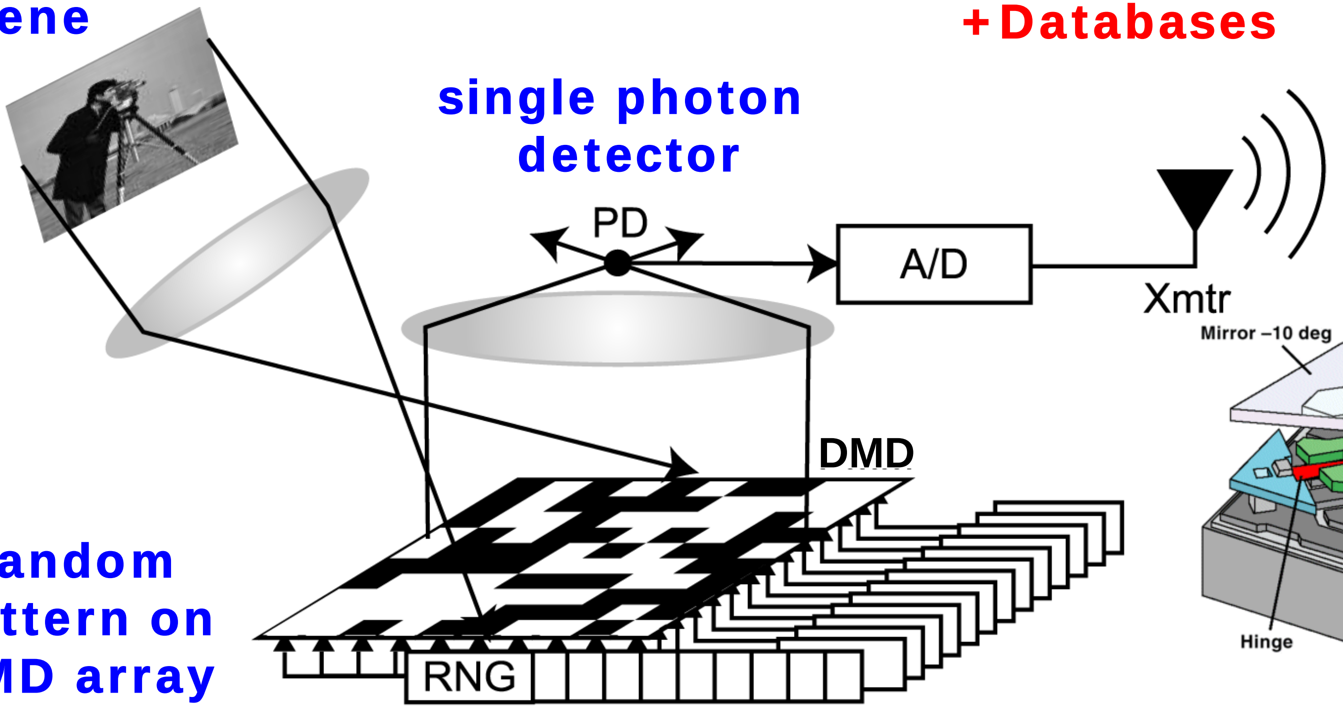
Sampling / sketching design

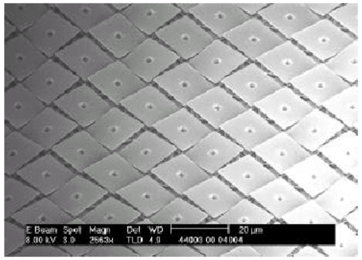
+Coding theory
+Theoretical computer science
+Learning theory
+Databases

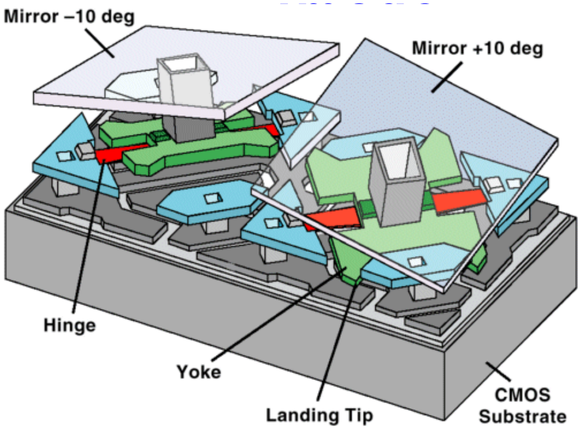
scene

single photon detector

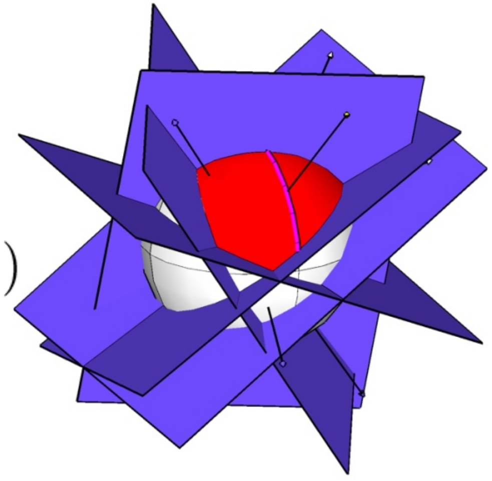
random pattern on DMD array

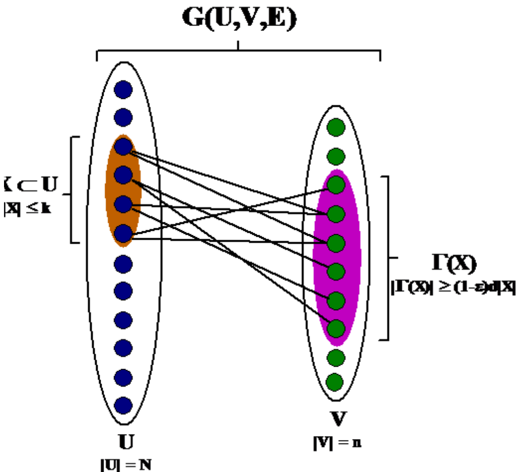






- Structured random matrices
- 1-bit CS $u = \text{sign}(\Phi x)$
- expanders & extractors

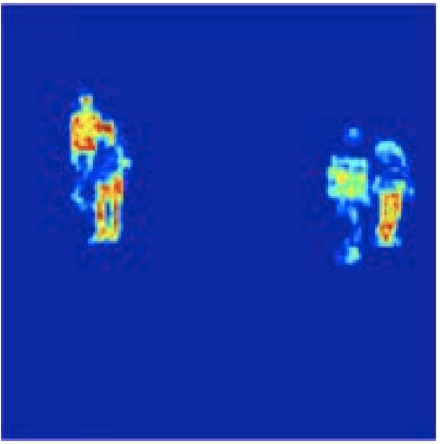


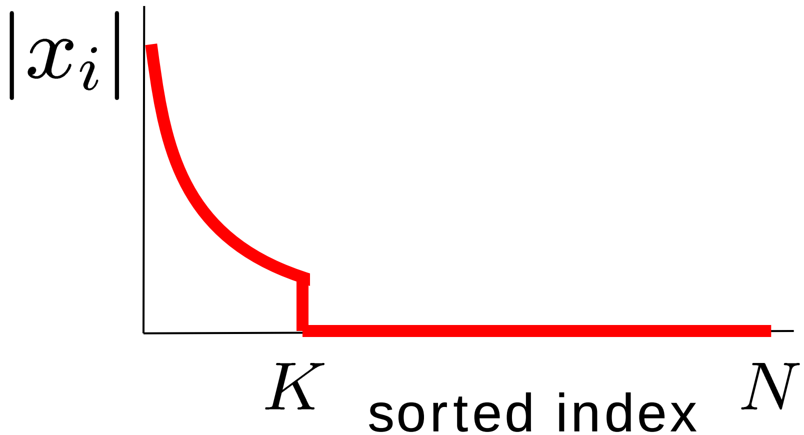


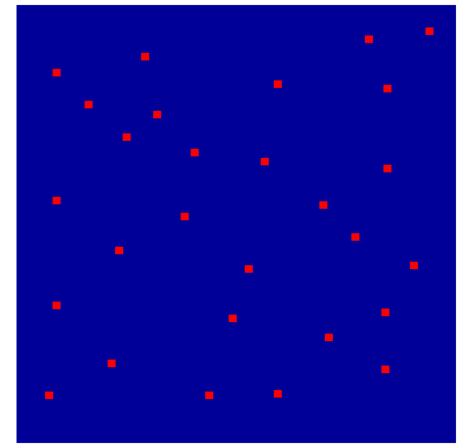
Structured recovery

+Theoretical computer science
+Learning theory
+Optimization
+Databases

- **Sparsity**





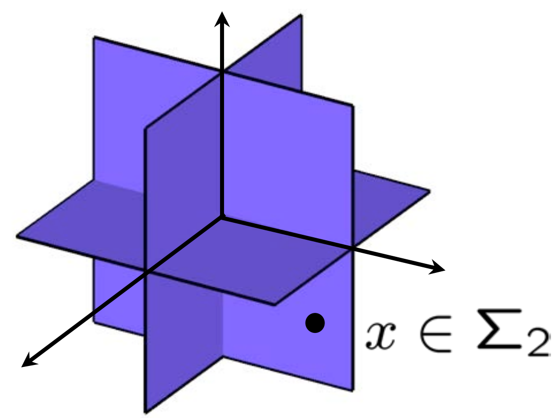


Sparse vector

only K out of N
coordinates nonzero

$$K \ll N$$

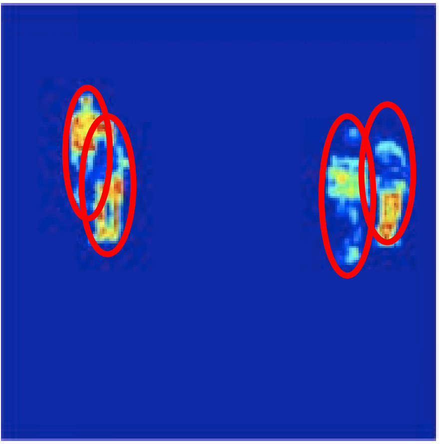
$$K = 2$$
$$\mathbf{R}^3$$

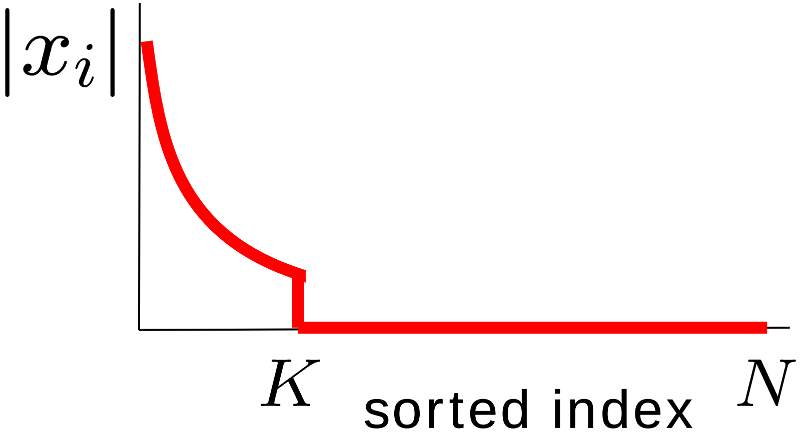


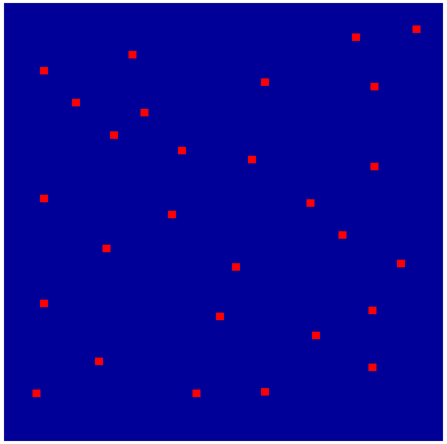
Structured recovery

+Theoretical computer science
+Learning theory
+Optimization
+Databases

- **Sparsity**



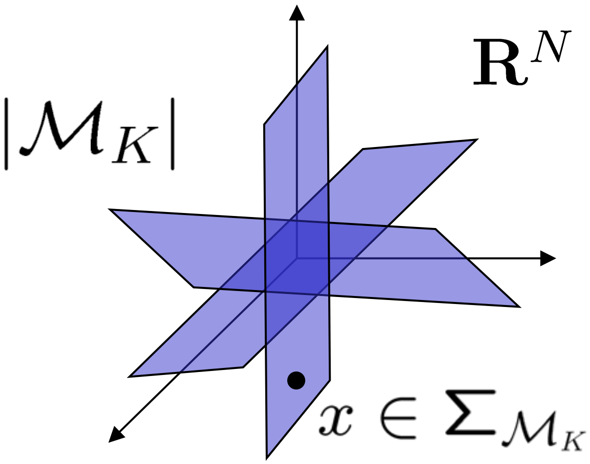




Structured sparse vector

only certain K out of N coordinates nonzero

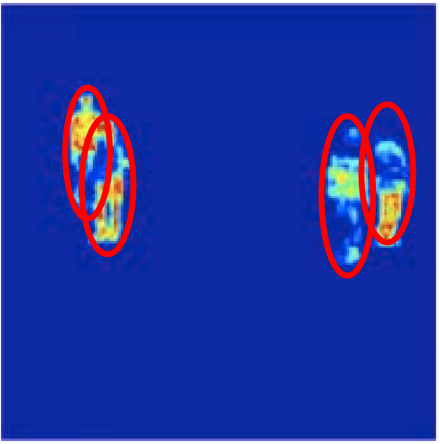
$K \ll N$



Structured recovery

+Theoretical computer science
+Learning theory
+Optimization
+Databases

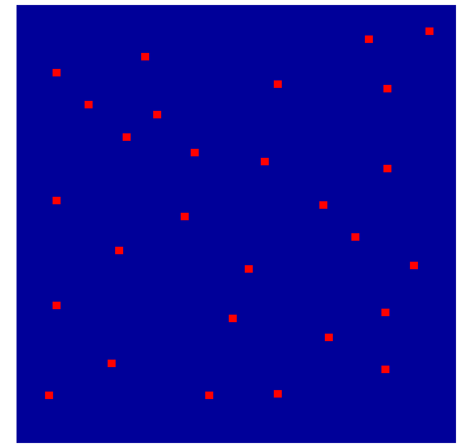
- **Structured sparsity**



\+ requires smaller sketches

\+ enhanced recovery

\+ faster recovery



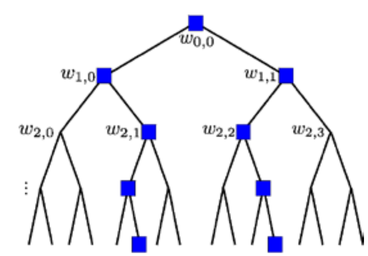
$$\mathrm{P}_{\Sigma_{\mathcal{M}}}(u;K) \in \arg\min_{x} \{\|x-u\| : x \in \Sigma_{\mathcal{M}_K}\}$$

support of the solution < > *modular approximation problem*
integer linear program

matroid structured sparse models

clustered / diversified sparsity models

tightly connected with max-cover, binpacking, knapsack problems

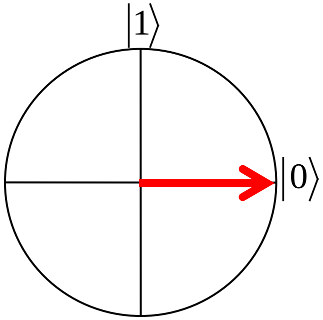


- Recovery with low-dimensional models, including low-rank...

Quantum tomography

+Theoretical computer science
+Databases
+Information theory
+Optimization

- **Quantum state estimation**



a state of n possibly-entangled qubits takes
~ 2^n bits to specify, even approximately

- **Recovery with rank and trace constraints**

with M=O(N)

1. *Create Pauli measurements (semi-random)*
2. *Estimate $Tr(\Phi_i \rho)$ for each $1 \leq i \leq M$*
3. *Find **any** "hypothesis state" σ st $Tr(\Phi_i \sigma) \approx Tr(\Phi_i \rho)$ for all $1 \leq i \leq M$*

- **Huge dimensional problem!**
  – (desperately) need scalable algorithms
  – also need theory for perfect density estimation

Learning theory and methods

+Learning theory
+Optimization
+Information theory
+Theoretical computer science

- A fundamental problem:

  given (y_i, x_i): $\mathbb{R} \times \mathbb{R}^d, i = 1, \ldots, m,$ learn a mapping $f\colon x \to y$

- Our interest < > non-parametric functions
  graphs (e.g., social networks)
  dictionary learning...

- Rigorous foundations < > sample complexity
  approximation guarantees
  tractability

- Key tools < > sparsity/low-rankness
  submodularity
  smoothness

Compressible priors

+Learning theory
+Statistics
+Information theory

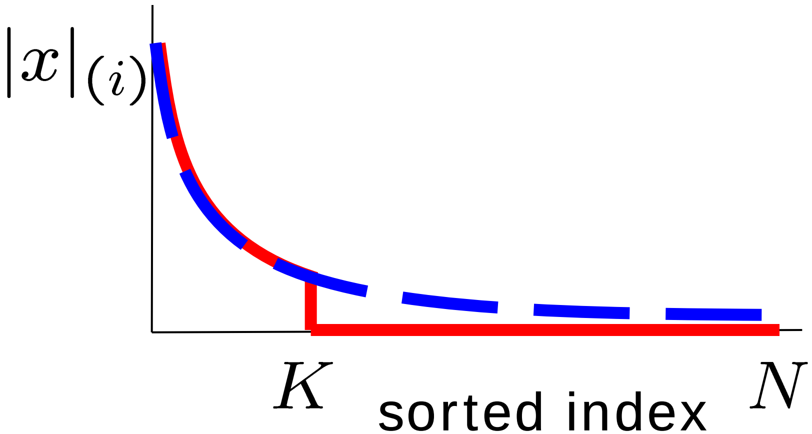
- **Goal:** seek distributions whose iid realizations $x_i \sim p(x)$ can be well-approximated as ***sparse***

Definition:

The PDF $p(x)$ is a *q-compressible prior* with parameters (ϵ, κ), when

$$\lim_{N\to\infty} \bar{\sigma}_{k_N}(x)_q \overset{a.s.}{\leq} \epsilon, \text{(a.s.: almost surely)};$$

for any sequence k_N such that $\lim_{N\to\infty} \inf \frac{k_N}{N} \geq \kappa$, where $\epsilon \ll 1$ and $\kappa \ll 1$.



relative k-term approximation:

$$\bar{\sigma}_k(x)_q = \frac{\sigma_k(x)_q}{\|x\|_q}$$

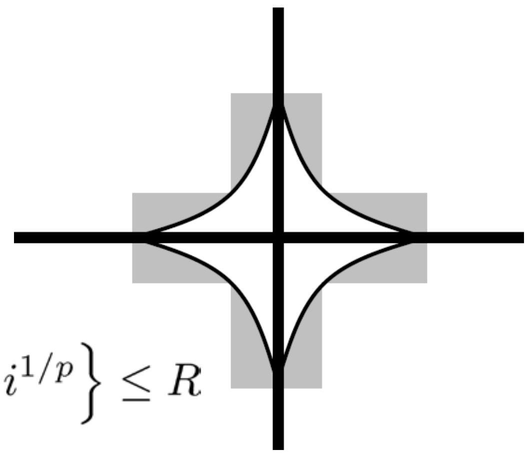
$$\sigma_k(x)_q := \inf_{\|u\|_0 \leq k} \|x - u\|_q$$

Compressible priors

+Learning theory
+Information theory

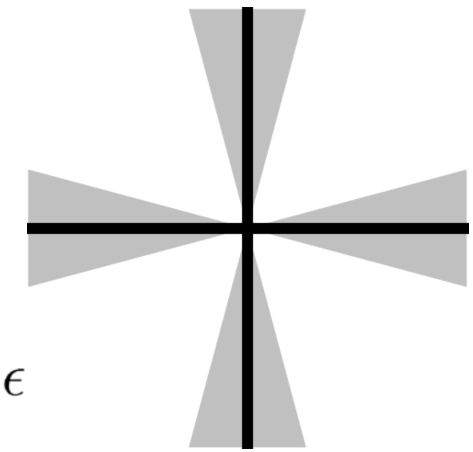
- **Goal:** seek distributions whose iid realizations can be well-approximated as ***sparse***

Classical:

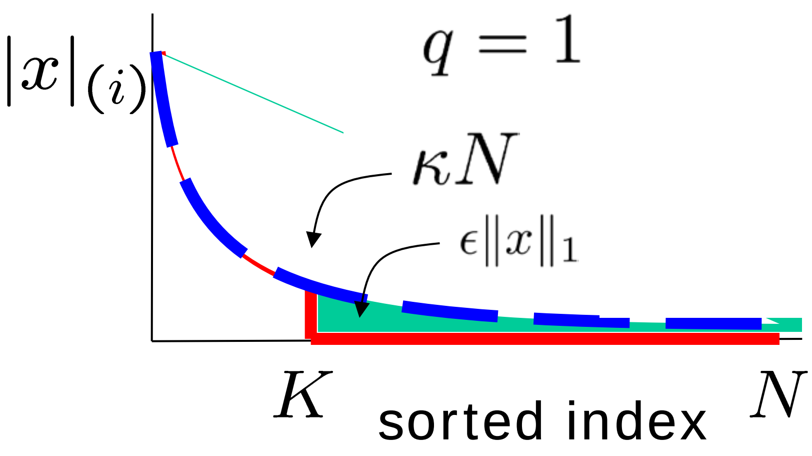


$$\|x\|_{w\ell^p} := \sup_i \left\{ |x|_{(i)} \cdot i^{1/p} \right\} \leq R$$

New:



$$\frac{\sigma_{\kappa N}(x)_q}{\|x\|_q} \leq \epsilon$$



Compressible priors

+Learning theory
+Statistics
+Information theory

- **Goal:** seek distributions whose iid realizations can be well-approximated as ***sparse***

- **Motivations:** deterministic embedding scaffold for the probabilistic view

  analytical proxies for sparse signals
  - learning (e.g., dim. reduced data)
  - algorithms (e.g., structured sparse)

  information theoretic (e.g., coding)

  lots of applications in vision, image understanding / analysis

References

M. Afonso, J. Bioucas-Dias, M. Figueiredo, “Fast image recovery using variable splitting and constrained optimization”, *IEEE Transactions on Image Processing, vol. 19, 2010.*

H. Attouch, J. Bolte, P. Redont, A. Soubeyran, “Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality”, *Math. Oper. Research*, 2010

O. Axelsson*, Iterative Solution Methods,* Cambridge University Press, 1996.

R. Baraniuk, V. Cevher, M. Duarte, C. Hegde, “Model-based compressive sensing”, *IEEE Transactions on Information Theory*, vol. 56, 2010.

R. Baraniuk, M. Davenport, R. de Vore, M. Wakin, “A Simple Proof of the Restricted Isometry Property for Random Matrices”, *Constructive Approximation*, 2008.

R. Baraniuk, V. Cevher, M. Wakin, “Low-dimensional models for dimensionality reduction and signal recovery: A geometric perspective”, *Proceedings of the IEEE*, vol. 98, 2010.

J. Barzilai and J. Borwein, “Two point step size gradient methods,” *IMA Journal of Numer. Anal., vol. 8, 1988.*

R. Basri, D. Jacobs, “Lambertian Reflectance and Linear Subspaces”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, 2003.

A. Beck, M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM Journal on Imaging Science, vol. 2,* 2009.

J. Bioucas-Dias, M. Figueiredo, “A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration,” *IEEE Transactions on Image Processing, vol. 16, 2007.*

P. Boufounos, R. Baraniuk, “1-bit compressed sensing”, *Proceedings of the Conference on Information Science and Systems*, Princeton, 2008.

C. Boutsidis, M. Mahoney, P. Drineas, ”An improved approximation algorithm for the column subset selection problem”, *Proc. 20th Annual ACM/SIAM Symposium on Discrete Algorithms,* New York*, NY,* 2008.

References

P. Bühlmann, S. van der Geer, *Statistics for High-Dimensional Data,* Springer, 2011.

E. Candès, "The restricted isometry property and its implications for compressed sensing", *Comptes Rendus Mathematique,* vol. 346, 2008.

E. Candès and B. Recht, "Exact matrix completion via convex optimization", *Foundations of Computational Mathematics*, vol. 9, 2009.

L. Carin, R. Baraniuk, V. Cevher, D. Dunson, M. Jordan, G. Sapiro, M. Wakin, "Learning low-dimensional signal models", *IEEE Signal Processing Magazine*, vol. 28, 2010.

V. Cevher, "An ALPS view of sparse recovery", *Proc. ICASSP,* 2011.

V. Chandrasekaran, B. Recht, P. Parrilo, A. Willsky, "The convex geometry of linear inverse problems", submitted, 2010.

R. Chartrand, W. Yin, "Iteratively reweighted algorithms for compressive sensing", *Proc. ICASSP*, 2008

S. Chen, D. Donoho, M. Saunders, "Atomic decomposition by Basis Pursuit", *SIAM Review*, vol. 43, 2001.

P. Combettes, V. Wajs, "Signal recovery by proximal forward-backward splitting", *SIAM Journal Multiscale Modeling and Simulation, vol. 4,* 2005.

J. Eckstein, D. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators", *Mathematical Programming, vol. 5, 1992.*

M. Figueiredo and J. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization", *IEEE Transactions on Image Processing,* vol. 19, 2010.

M. Figueiredo, R. Nowak, S. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems", *IEEE Journal of Selected Topics in Signal Processing,* vol. 1, 2007.

References

D. Gabay, B. Mercier, “A dual algorithm for the solution of nonlinear variational problems via finite-element approximations”, *Computers and Mathematics with Application, vol. 2, 1976.*

R. Glowinski, A. Marroco, “Sur l’approximation, par elements finis d’ordre un, et la resolution, par penalisation-dualité d’une classe de problemes de Dirichlet non lineares,” *Rev. Française d’Automatique, 1975.*

Y. Gordon, “On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$”, in *Geometric Aspects of Functional Analysis,* Springer, 1988.

E. Hale, W. Yin, Y. Zhang, “Fixed-point continuation for l1-minimization: Methodology and convergence”, *SIAM Journal on Optimization*, vol. 19, 2008.

N. Halko, P.-G. Martinsson, J. Tropp, “Finding structure with randomness: stochastic algorithms for constructing approximate matrix decompositions”, *SIAM Review*, vol. 53, 2011.

C. Hegde, M. Duarte, V. Cevher, “Compressive sensing recovery of spike trains using a structured sparsity model”, *Proceedings of SPARS'09*, Saint-Malo, France, 2009.

M. Hestenes, “Multiplier and gradient methods”, *Journal of Optimazion Theory andApplications, vol. 4, 1969.*

A. Kyrillidis, V. Cevher, “Recipes for hard thresholding methods”, Tech. Rep., EPFL, 2011.

J. Lee, V. Mirrokni, V. Nagarajan, M. Sviridenko, “Non-monotone submodular maximization under matroid and knapsack constraints”, *Proc. 41st Annual ACM Symposium on Theory of Computing, Bethesda, MD,* 2009.

A. Lewis, J. Malick, “Alternating projections on manifolds”, *Math. of Operations Research*, vol. 33, 2008.

D. Lorenz, “Constructing test instances for basis pursuit denoising”, submitted, 2011.

N. Meinshausen, P. Bühlmann, “High-dimensional graphs and variable selection with the lasso”, *The Annals of Statistics*, vol. 34, pp. 1436-1462, 2006.

J.-J. Moreau, “Proximité et dualité dans un espace hilbertien,” *Bull. Soc. Mathematiques de France, vol. 93, 1965.*

References

G. Nemhauser, L. Wolsey, *Integer and combinatorial optimization*, Wiley, 1988.

S. Osher, M. Burger, D. Goldfarb, J. Xu, W. Yin, "An iterative regularization method for total variation-based image restoration", *SIAM Journal on Multiscale Modeling and Simulation*, vol. 4, 2005.

Y. Plan, "Compressed sensing, sparse approximation, low-rank matrix estimation", PhD Thesis, Caltech, 2011

M. Powell, "A method for nonlinear constraints in minimization problems", in *Optimization,* Academic Press*, 1969.*

H. Raguet, J. Fadili, G. Peyré, "Generalized Forward-Backward splitting", Tech. report, Hal-00613637, 2011.

S. Setzer, G. Steidl, T. Teuber, "Deblurring Poissonian images by split Bregman techniques," *Journal of Visual Communication and Image Representation, 2010.*

K.-C. Toh , S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized least squares Problems", *Pacific Journal of Optimization*, vol. 6, 2010.

A. Waters, A. Sankaranarayanan, R. Baraniuk, "SpaRCS: recovering low-rank and sparse matrices from compressive measurements", *Neural Information Processing Systems,* 2011.

S. Wright, R. Nowak, M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Transactions on Signal Processing, vol. 57,* 2009.

W. Yin, S. Osher, D. Goldfarb, J. Darbon, "Bregman iterative algorithms for l1-minimization with applications to compressed sensing", *SIAM Journal on Imaging Science, vol. 1, 2008.*

T. Zhou, D. Tao, "Godec: randomized low-rank & sparse matrix decomposition in noisy case," *Proc. International Conference on Machine Learning*, Bellevue, WA, 2011.