

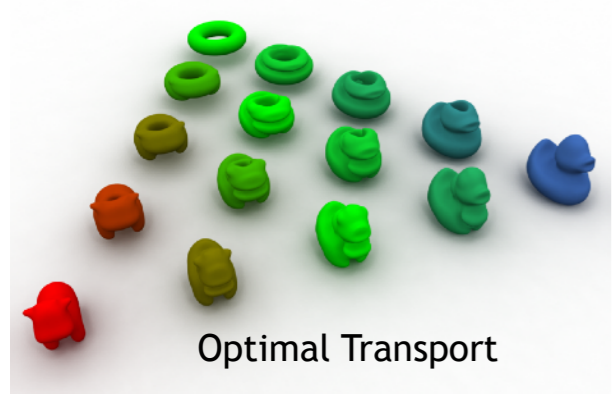
Mathematical Coffees



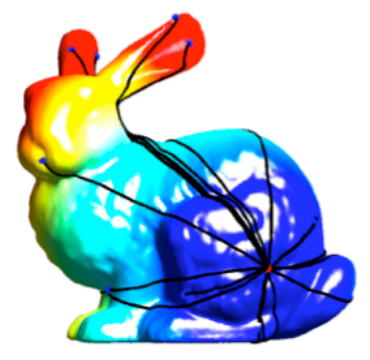
FSMP
Fondation Sciences
Mathématiques de Paris

Huawei-FSMP joint seminars
<https://mathematical-coffees.github.io>

Organized by: Mérouane Debbah & Gabriel Peyré



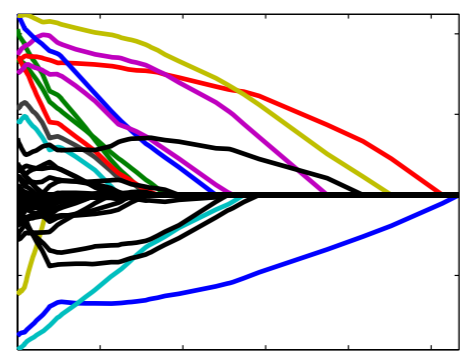
Optimal Transport



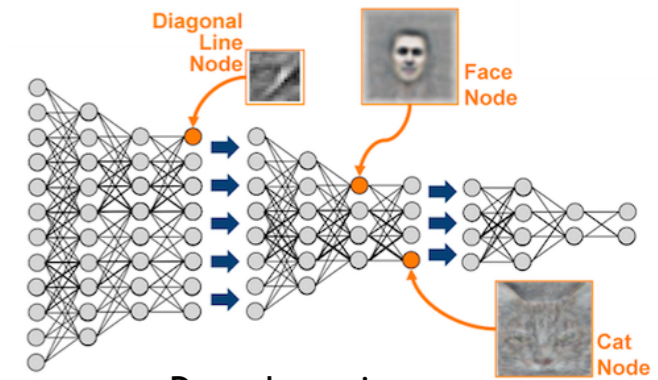
Geodesics



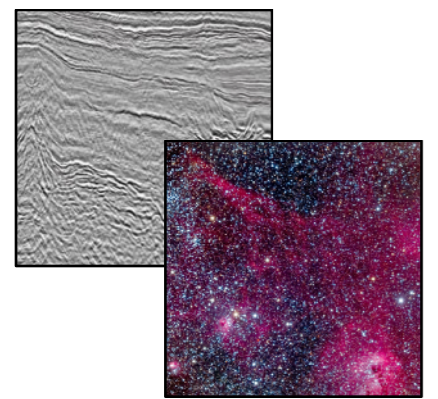
Meshes



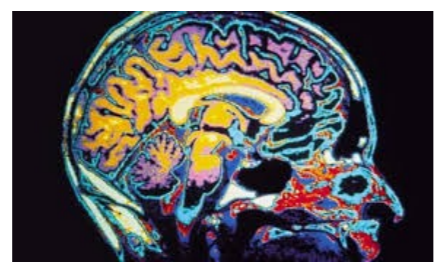
Optimization



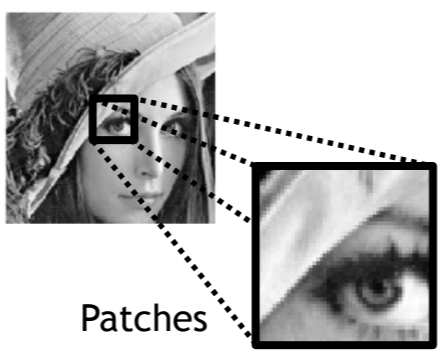
Deep Learning



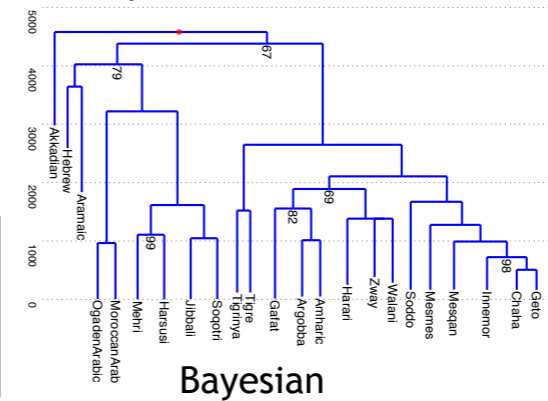
Sparsity



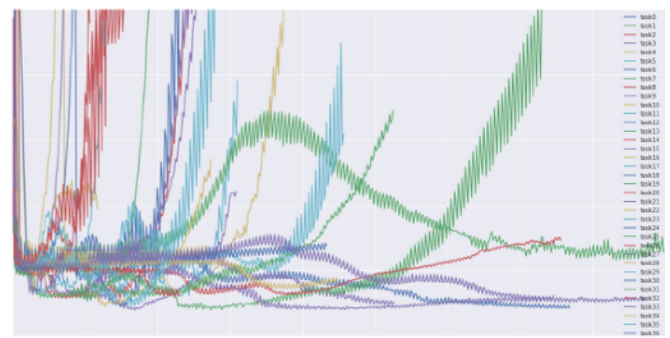
Neuro-imaging



Patches



Bayesian



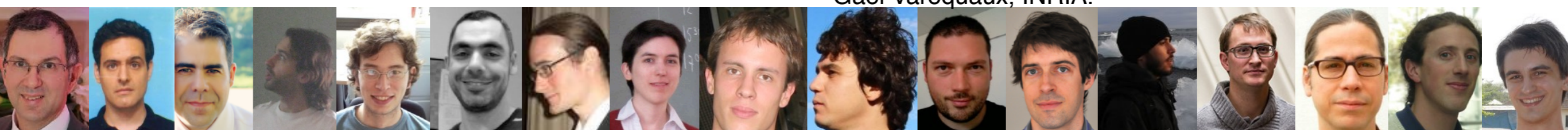
Parallel/Stochastic

Alexandre Allauzen, Paris-Sud.
Pierre Alliez, INRIA.
Guillaume Charpiat, INRIA.
Emilie Chouzenoux, Paris-Est.

Nicolas Courty, IRISA.
Laurent Cohen, CNRS Dauphine.
Marco Cuturi, ENSAE.
Julie Delon, Paris 5.

Fabian Pedregosa, INRIA.
Guillaume Lécué, CNRS ENSAE
Julien Tierny, CNRS and P6.
Robin Ryder, Paris-Dauphine.
Gael Varoquaux, INRIA.

Jalal Fadili, ENSICAen.
Alexandre Gramfort, INRIA.
Matthieu Kowalski, Supelec.
Jean-Marie Mirebeau, CNRS,P-Sud.

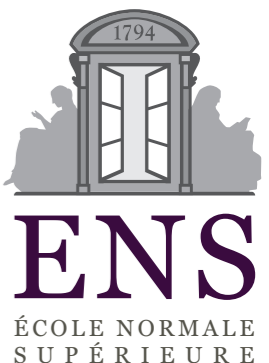


Curses and Blessings of High Dimension

Gabriel Peyré



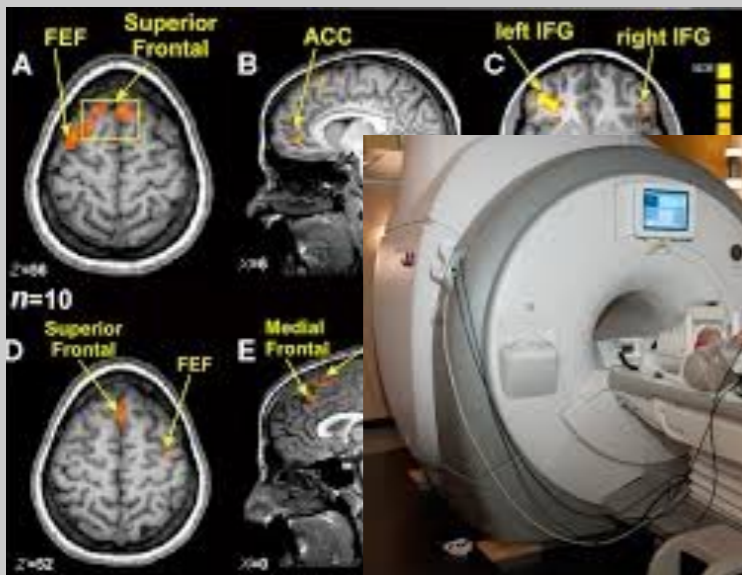
www.numerical-tours.com



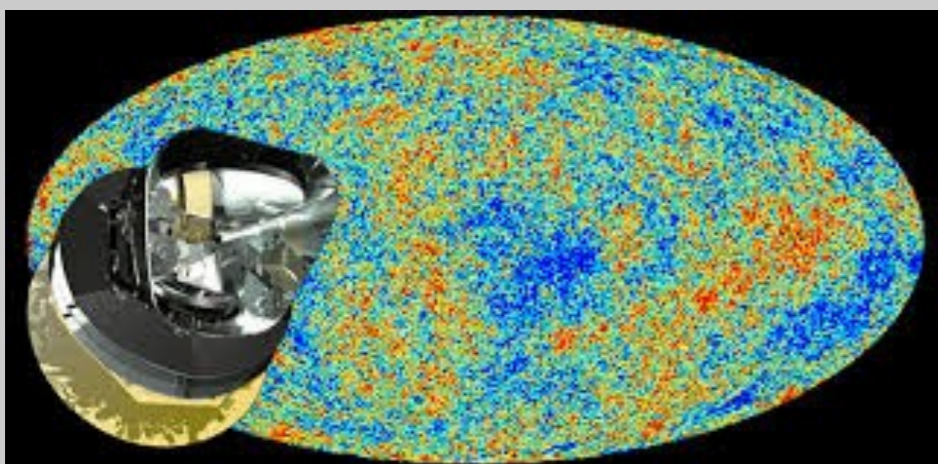
Big Datasets (large n)



Giga pixels camera



fMRI imaging



Planck mission

Imaging sciences

Machine learning



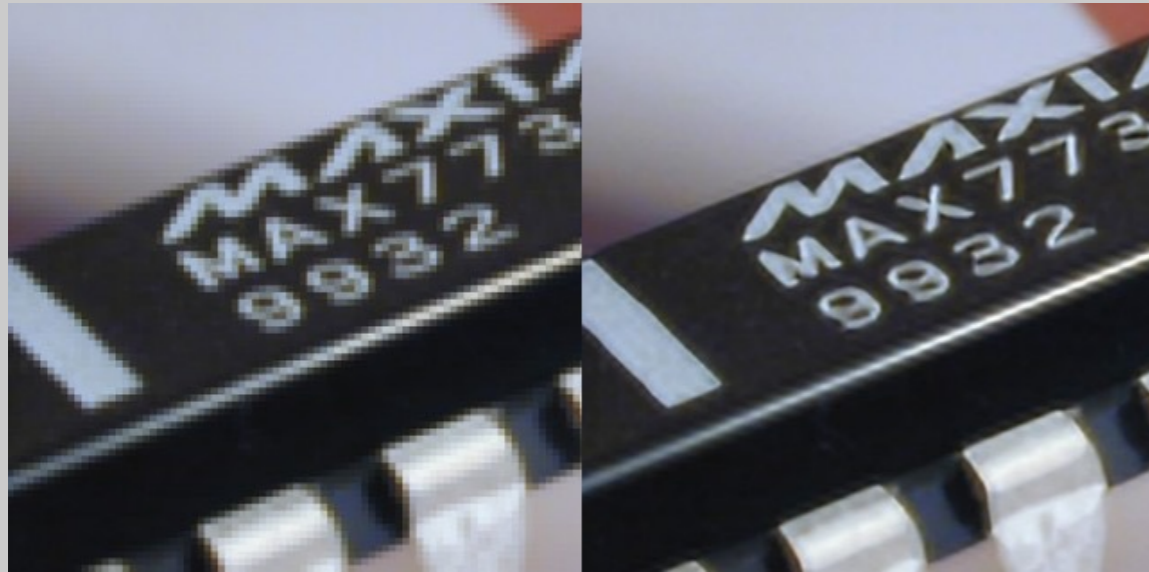
DNA microarrays



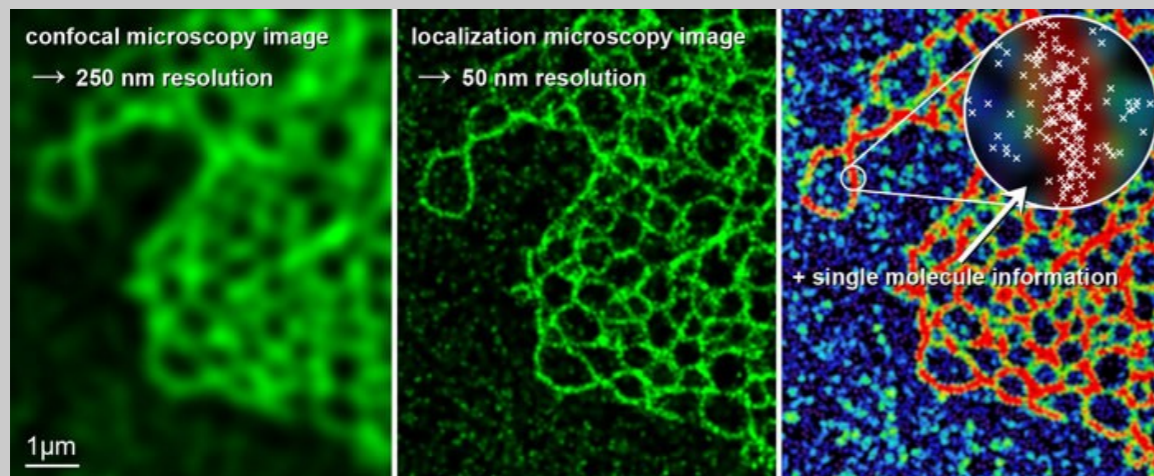
Network monitoring

Big Models (large p)

Super-resolution



Single molecule imaging



Imaging sciences

Machine learning

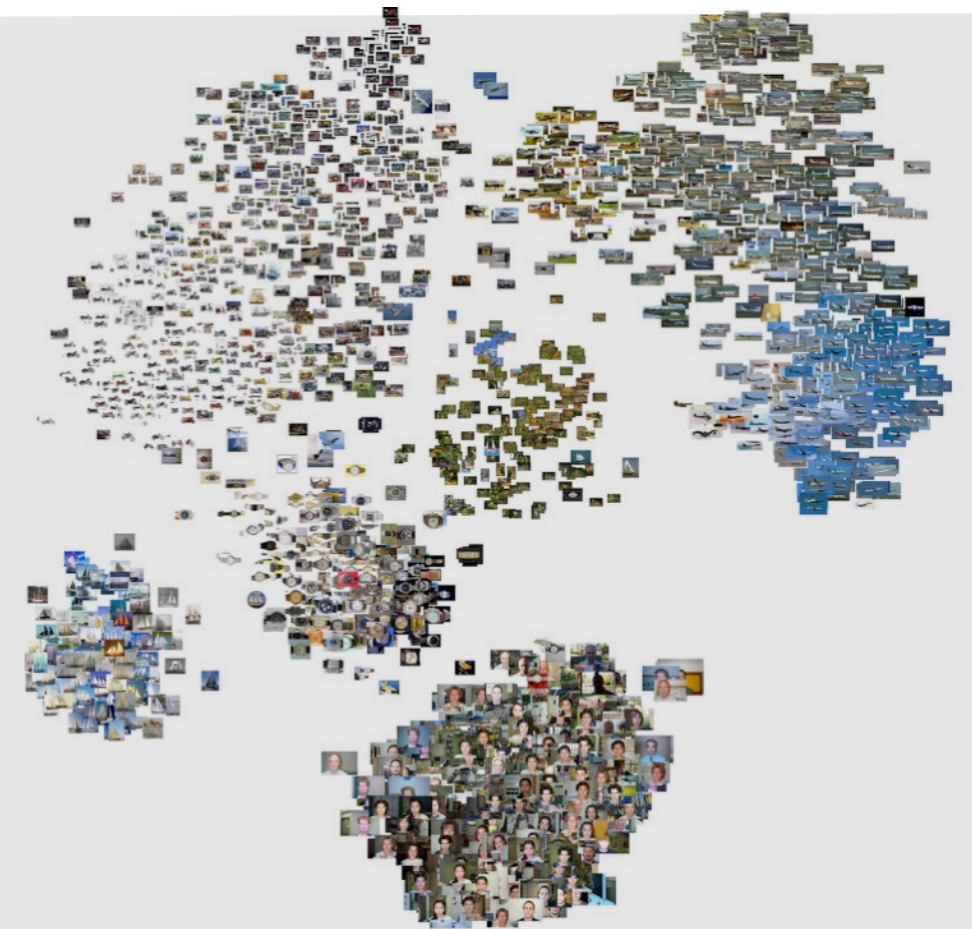
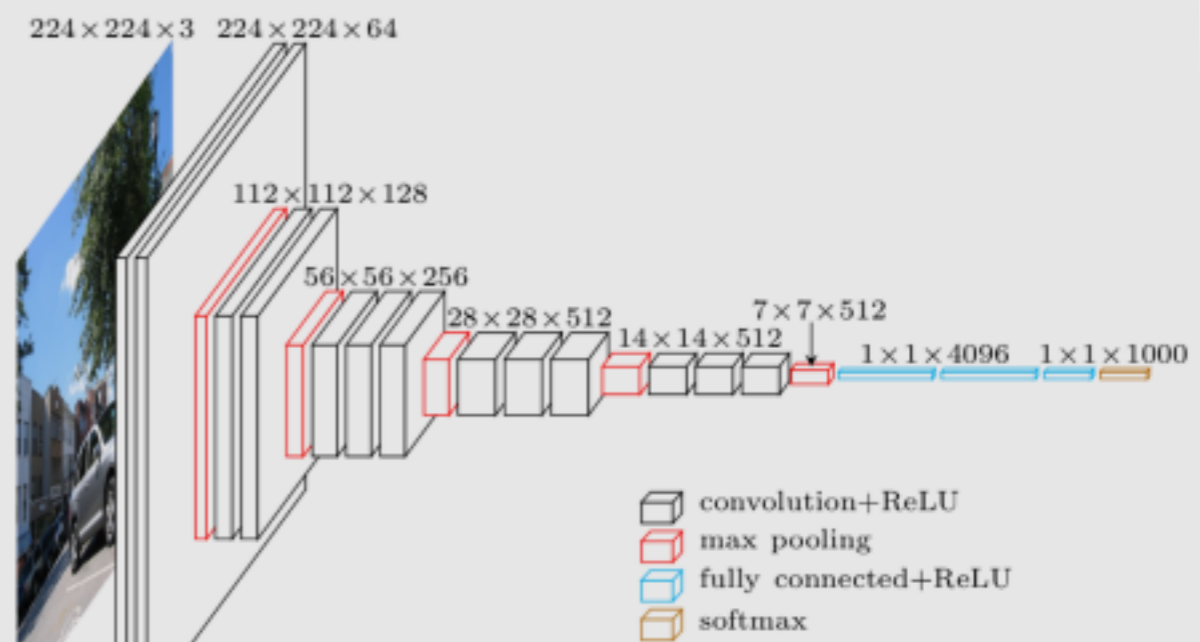


Image classification



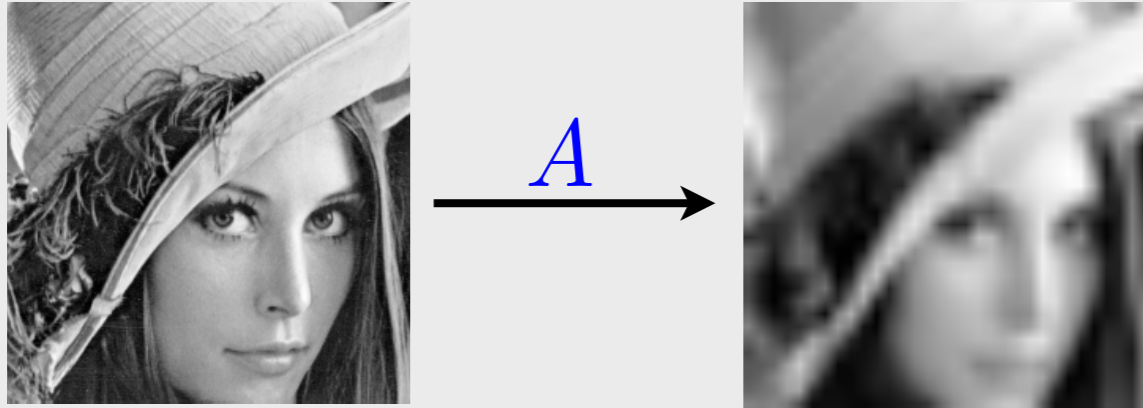
Very deep learning

Linear Models

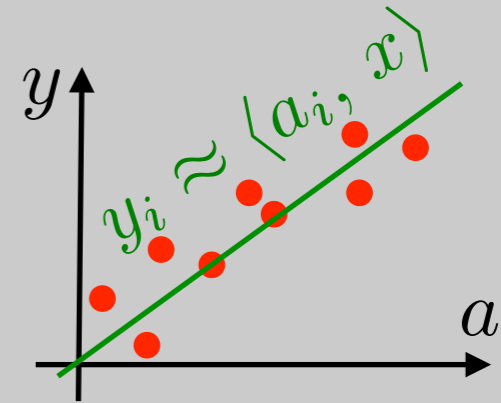
Solving $y \underset{\approx}{=} Ax \in \mathbb{R}^n$

$$A \in \mathbb{R}^{n \times p}$$

Imaging sciences:



Machine learning:
Observations $(a_i, y_i)_{i=1}^p$,

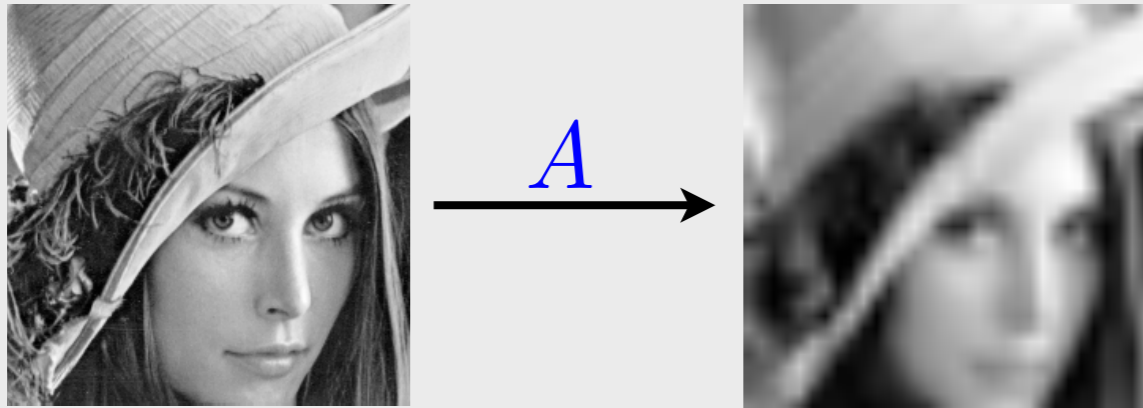


Linear Models

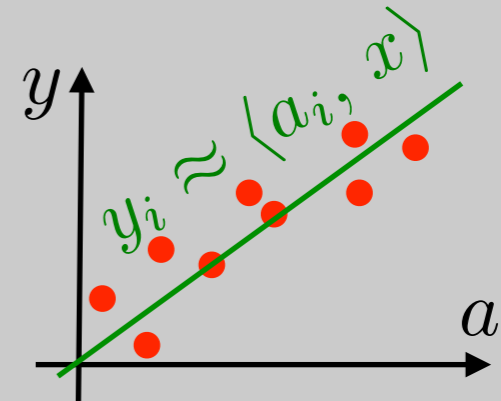
Solving $y \approx Ax \in \mathbb{R}^n$

$$A \in \mathbb{R}^{n \times p}$$

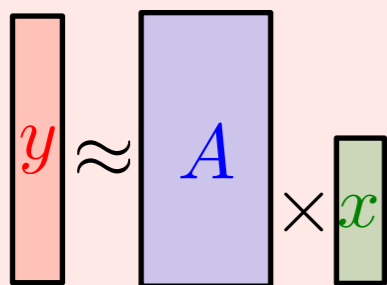
Imaging sciences:



Machine learning:
Observations $(a_i, y_i)_{i=1}^p$,

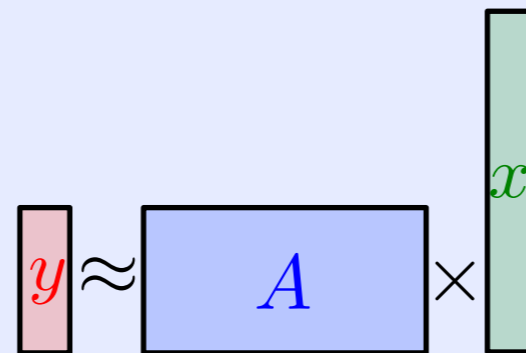


Over-determined ($n > p$)

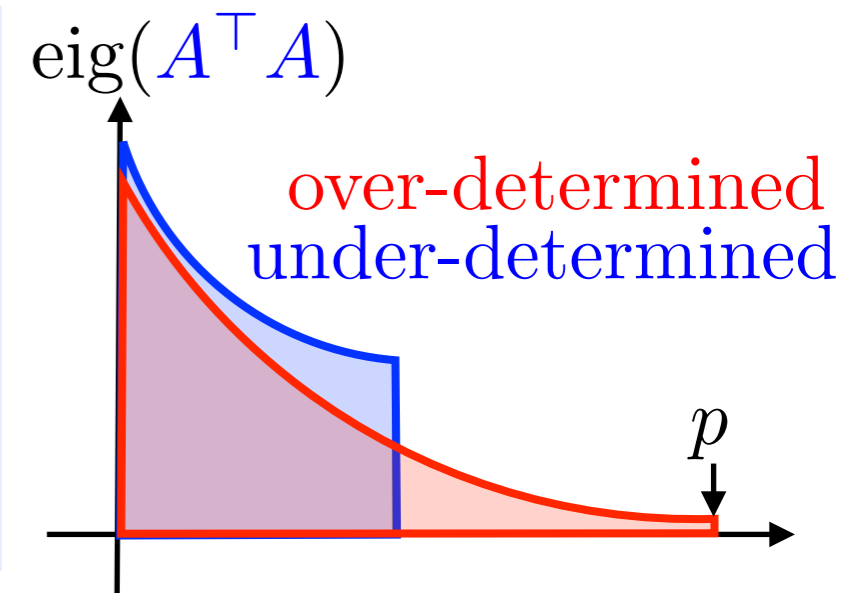


A diagram representing an over-determined system. It shows a tall red vertical bar labeled y , followed by an approximation symbol \approx , a purple square labeled A , a multiplication symbol \times , and a short green vertical bar labeled x .

Under-determined ($n < p$)



A diagram representing an under-determined system. It shows a short red vertical bar labeled y , followed by an approximation symbol \approx , a wide blue rectangle labeled A , a multiplication symbol \times , and a tall green vertical bar labeled x .

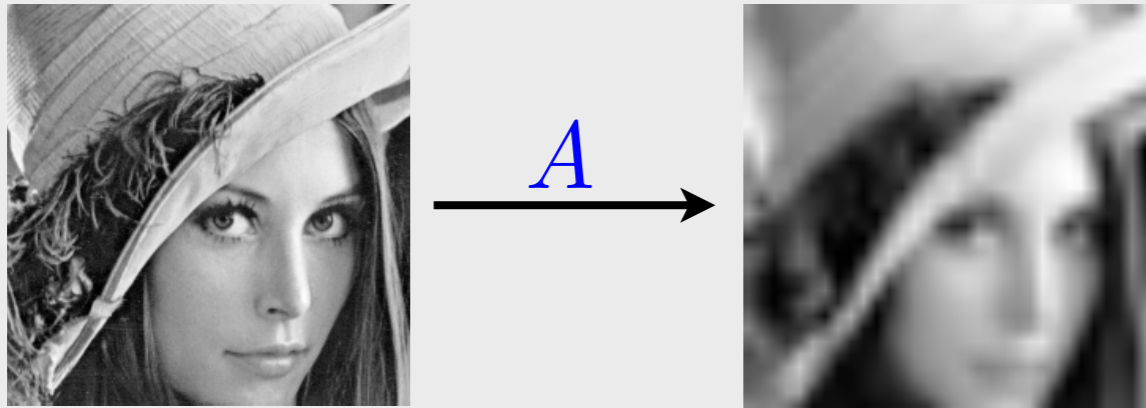


Linear Models

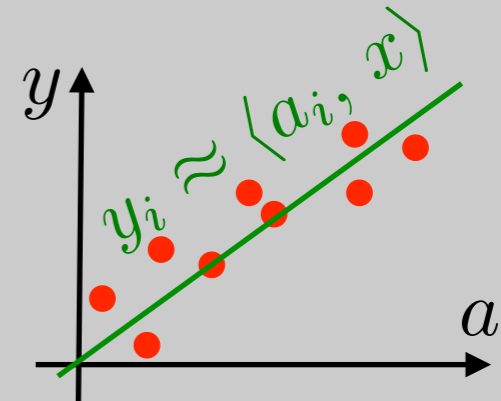
Solving $y \approx Ax \in \mathbb{R}^n$

$$A \in \mathbb{R}^{n \times p}$$

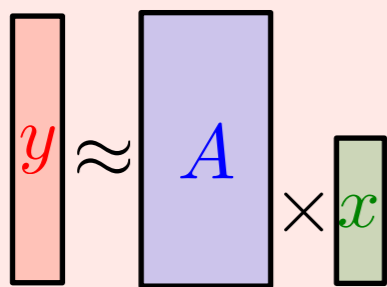
Imaging sciences:



Machine learning:
Observations $(a_i, y_i)_{i=1}^p$,

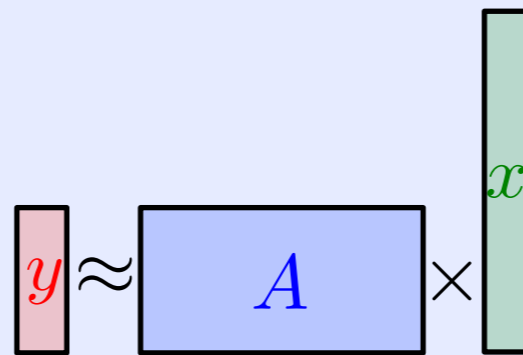


Over-determined ($n > p$)

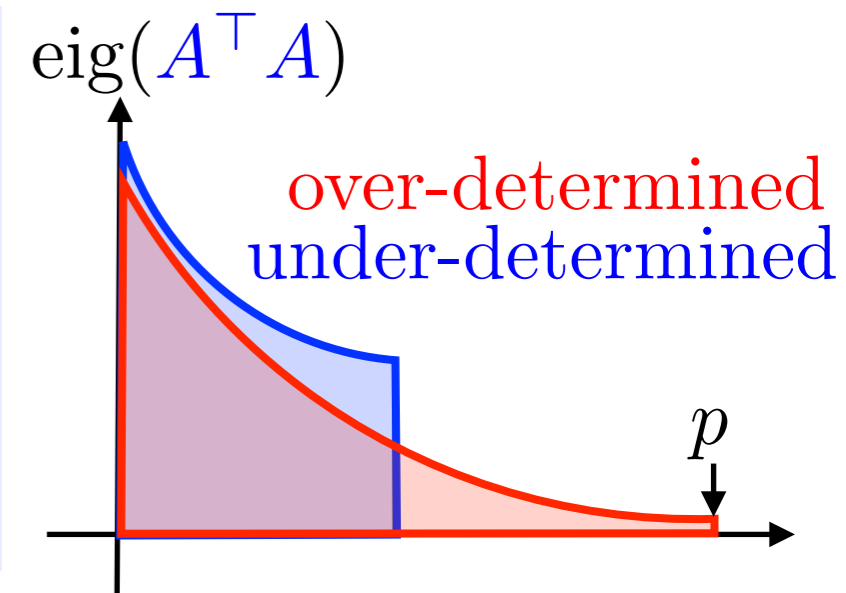
$$y \approx Ax$$


A diagram illustrating an over-determined system. A tall red vertical bar labeled y is on the left, followed by a blue square labeled A , and a shorter green vertical bar labeled x on the right. The equation $y \approx Ax$ is written between them.

Under-determined ($n < p$)

$$y \approx Ax$$


A diagram illustrating an under-determined system. A shorter red vertical bar labeled y is on the left, followed by a wide blue rectangle labeled A , and a tall green vertical bar labeled x on the right. The equation $y \approx Ax$ is written between them.



Curse: Ill-posed, noisy, large size (n, p) .

Blessing: unreasonable effectiveness of regularization in high dimension.

Algorithms for large (n,p)

Regularized least square / empirical risk minimization:

$$\min_f \|Ax - y\|^2 + \lambda \|x\|^2$$

$$x = \underbrace{(A^\top A + \lambda \text{Id}_p)^{-1} A^\top y}_{\substack{\text{If } n > p \\ \text{(over-determined)}}} = \underbrace{A^\top (AA^\top + \lambda \text{Id}_n)^{-1} y}_{\substack{\text{If } n < p \\ \text{(under-determined)}}$$

(kernel methods:
 $p = +\infty$)

Algorithms for large (n,p)

Regularized least square / empirical risk minimization:

$$\min_f \|Ax - y\|^2 + \lambda \|x\|^2$$

$$x = \underbrace{(A^\top A + \lambda \text{Id}_p)^{-1} A^\top y}_{\substack{\text{If } n > p \\ \text{(over-determined)}}} = \underbrace{A^\top (AA^\top + \lambda \text{Id}_n)^{-1} y}_{\substack{\text{If } n < p \\ \text{(under-determined)}}$$

(kernel methods:
 $p = +\infty$)

Large but finite (n, p): use first order methods.

Gradient descent, CG, BFGS, proximal splittings.

→ $O(np)$ or even $O(p)$ cost per iterate.

→ Extends to non-smooth regularization (e.g. ℓ^1).

Algorithms for large (n,p)

Regularized least square / empirical risk minimization:

$$\min_f \|Ax - y\|^2 + \lambda \|x\|^2$$

$$x = \underbrace{(A^\top A + \lambda \text{Id}_p)^{-1} A^\top y}_{\substack{\text{If } n > p \\ \text{(over-determined)}}} = \underbrace{A^\top (AA^\top + \lambda \text{Id}_n)^{-1} y}_{\substack{\text{If } n < p \\ \text{(under-determined)}}$$

(kernel methods:
 $p = +\infty$)

Large but finite (n, p): use first order methods.

Gradient descent, CG, BFGS, proximal splittings.

→ $O(np)$ or even $O(p)$ cost per iterate.

→ Extends to non-smooth regularization (e.g. ℓ^1).

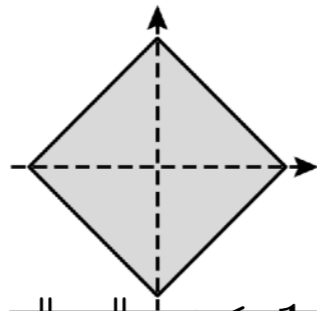
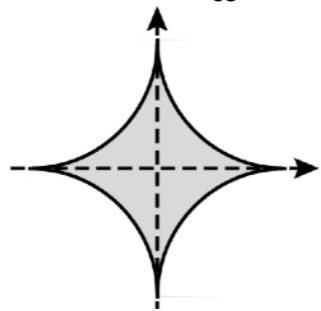
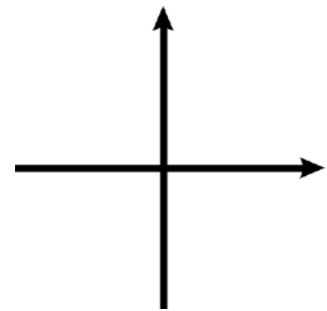
Very large or infinite n: use stochastic descent methods.

Draw (y_i, a_i) at random, then $x \leftarrow (1 - \tau_k \lambda)x - \tau_k (\langle a_i, x \rangle - y_i) a_i$
decays to 0

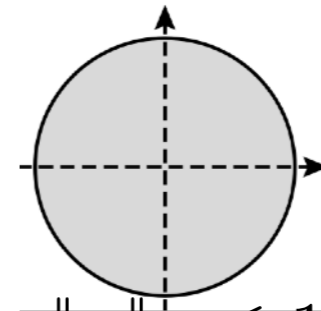
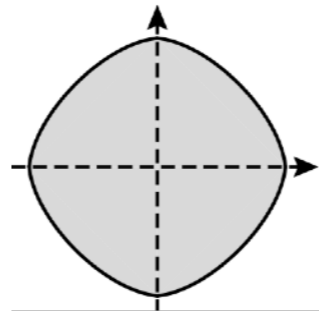
L1 and Dimensionality Reduction

Sparsity / model selection: replace $\|x\|^2$ by $\|x\|_1$.

$$\min_x \|Ax - y\|^2 + \lambda \|x\|_1$$



$$\|x\|_1 \leq 1$$



$$\|x\|_2 \leq 1$$



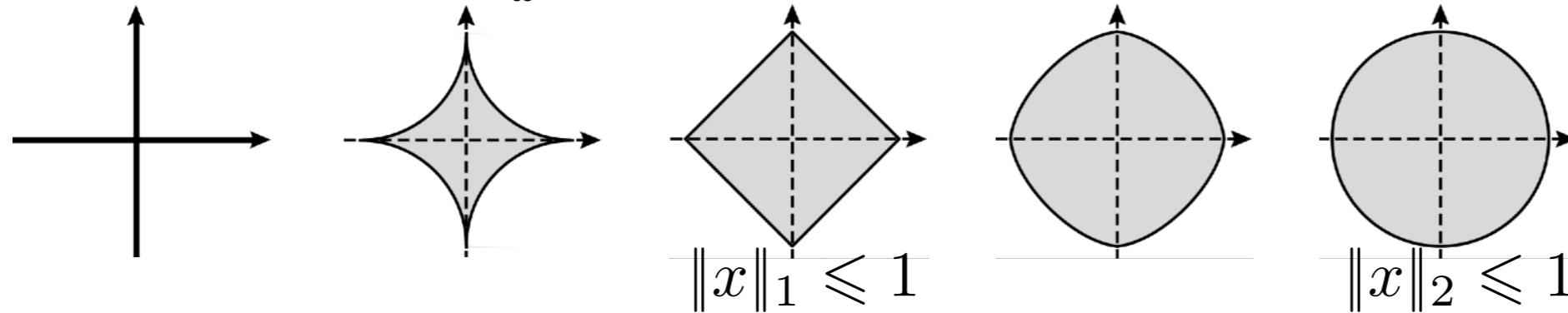
misleading
in high-dimension

→ Better model in imaging sciences. → Support recovery with very large p .

L1 and Dimensionality Reduction

Sparsity / model selection: replace $\|x\|^2$ by $\|x\|_1$.

$$\min_x \|Ax - y\|^2 + \lambda \|x\|_1$$

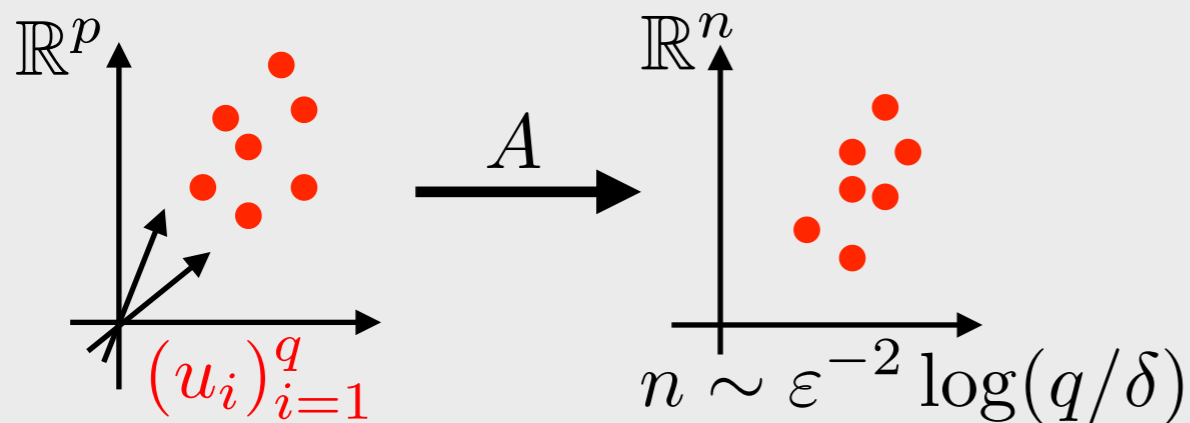


misleading
in high-dimension

→ Better model in imaging sciences. → Support recovery with very large p .

“Optimal” setting: choose $A \in \mathbb{R}^{n \times p}$ random.

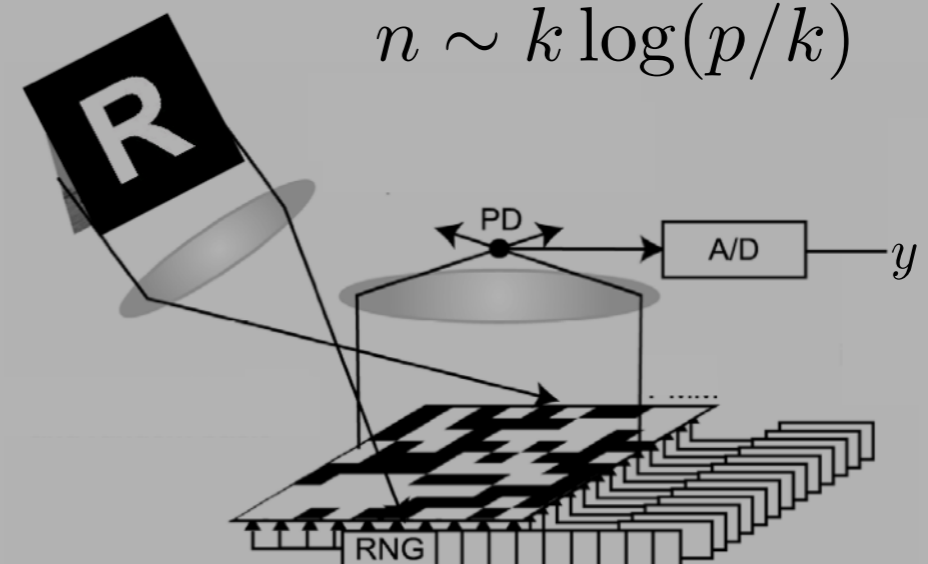
Johnson-Lindenstrauss lemma



$$\frac{\|A(u_i - u_j)\|_{\mathbb{R}^n}^2}{\|u_i - u_j\|_{\mathbb{R}^p}^2} \in [(1 - \varepsilon), (1 + \varepsilon)]$$

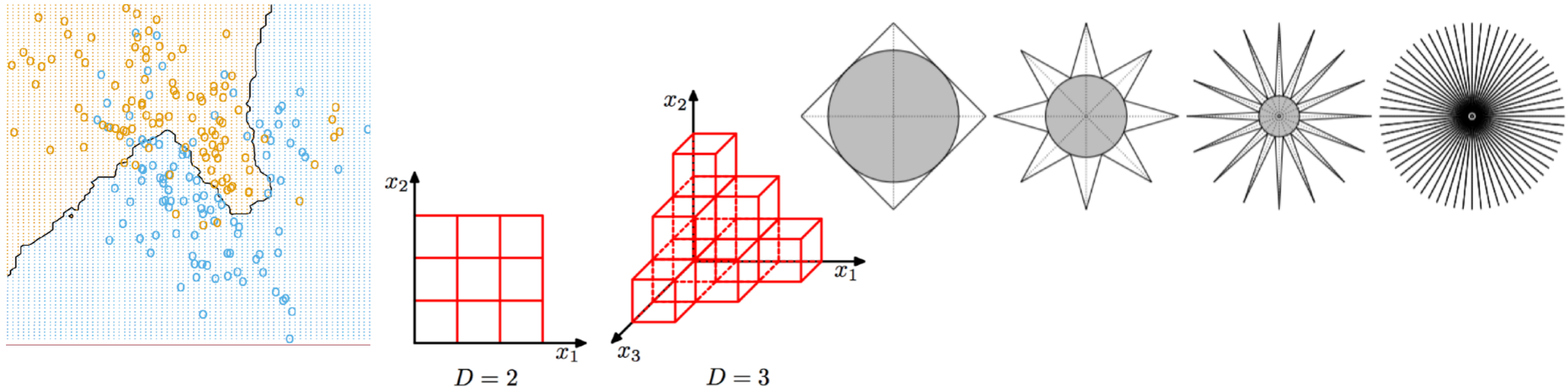
Compressed sensing
Perfect recovery of k -sparse input

$$n \sim k \log(p/k)$$



What's Next

Julie Delon: not so intuitive phenomena in high dimension.



Jalal Fadili: model selection in high-dimension.

