Applied Maths meets Data Sciences

Gabriel Peyré

www.gpeyre.com
www.numerical-tours.com

Overview

- FSMP
- Carnot SMILE
- Data Science in Paris and in France
- Huawei-FSMP Mathematical Coffees
- My Past and Present Projects
- Computational Optimal Transport

Fondation Sciences Mathématiques de Paris (FSMP)

President: Jean-Charles Pomerol Director: Emmanuel Trélat

- Creation in december 2006: RTRA in Mathematical Sciences
  $\longrightarrow$ Fondation de Coopération Scientifique (FCS)
- Founder members:
  CNRS, ENS, Paris 6, Paris 7 (+ 4 Chairs at Collège de France)
  Partners:
  Dauphine, Inria, Paris 5, Paris 13, Paris 1, Mines, Obs. Paris, EHESS
- 1800 researchers among which 900 permanents
- Financial supports:
  - Regional: DIM project "Math Innov"
  - National: LabEx SMP
  - International: Cofund "MathInParis"

  + Contributions of founders and partners, Ville de Paris, financial interest earnings, patronage, industrials.

FSMP
Fondation Sciences Mathématiques de Paris

Our programs

4 core programs

- PGSM (Paris Graduate School of Mathematics):
  funding of scholarships and training of international **master** students
  (contacts in embassies around the world)
- Doctoral program: funding and training of **PhD** students
- Post-doc program:
  funding and training of international **post-doctorate** researchers
- Excellence Chairs:
  funding of **Chairs** for international prestigious researchers

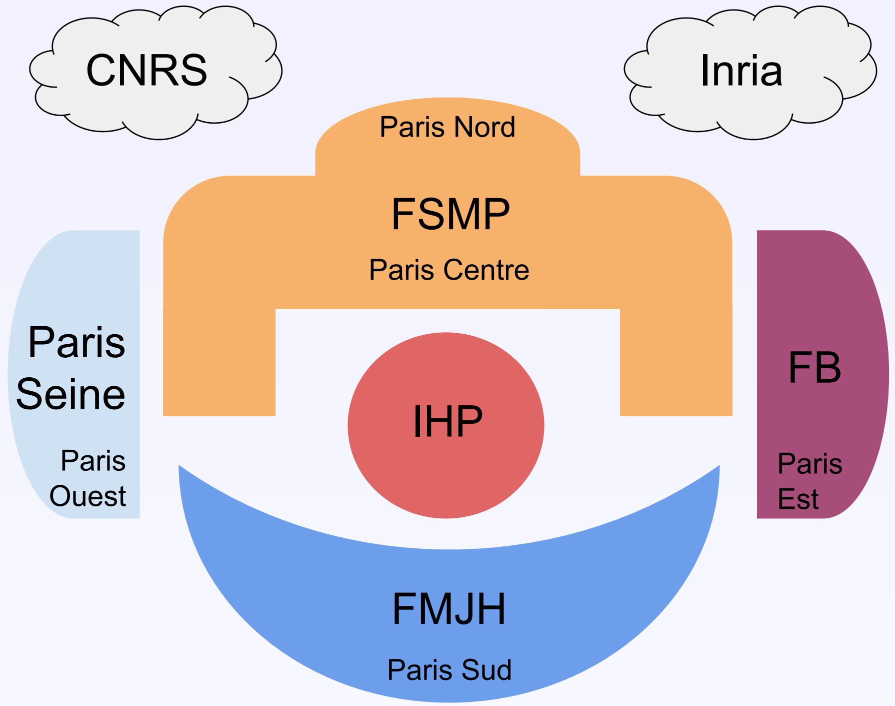
+ many other "small" programs

- FSMP is a tool at the service of the departments and teams of its perimeter and beyond (regional, national, international operations).
- Coordination of concerted actions with mathematics LabEx in France.

FSMP in the Ile-de-France network

- FSMP = manager and federating operator of the DIM project "Math Innov" (Région IdF, 2017–2020)

FSMP $\longrightarrow$ single portal to mathematics in IdF (3000 researchers)



- Many joint operations:
  - Co-funding of Chairs, PhDs, post-docs (also with IdEx).
  - With IHP: thematic trimesters, workshops, communication, discussion-evenings with our industrial partners, actions for young people, etc.
  - Application of Paris to International Congress of Mathematics 2022.

Maths and companies

- Actions for students:
  - Challenge Data → machine learning competitions (700 students in 2016)
    → incentive and incubation of start-ups
  - Industrial PhDs and post-docs in the new DIM project.
  - Google PhD Fellowships.
  - Forum Emploi Maths (many partners)
  - Training sessions for job dating (PhDs, post-docs).

- Workshops, meetings:
  - Mathématiques Oxygène du Numérique (FSMP, IHP, AMIES, oct. 2016).
  - Horizon Maths: **Huawei, fall 2017 (with IHP)**.
  - Discussion-evenings at IHP (for example: J.-B. Rudelle, CRITEO)

- Research collaborations:
  - Institut Carnot *SMILES* (Sciences Mathématiques pour l'Innovation: Label d'Excellence Stratégique)
  - Expertise, consulting (Airbus, Air Liquide, **Huawei**, LVMH, ...).

Conclusion

Regional, national and international impact of FSMP

- Federating role and coordination of mathematical sciences
- Visibility of mathematical sciences for other thematics, society, industrial partners, at the international level
  $\Rightarrow$ role of incubator
- Excellence international programs: attractivity
- Multiplying effect: Initial grant by government: 15 M€
  Cumulated budget over 10 ans: 30 M€

We promote excellence of mathematics at the highest international level.

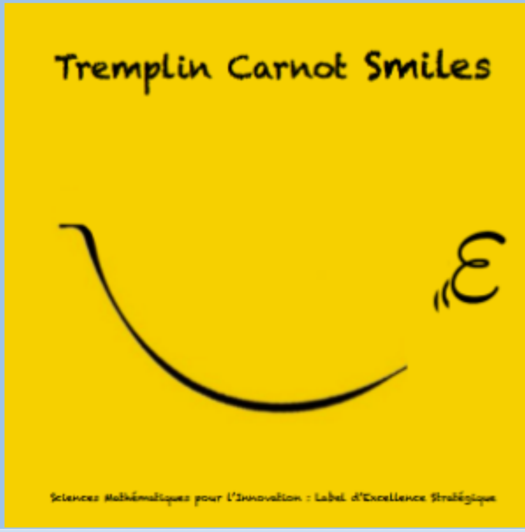
Overview

- FSMP
- Carnot SMILE
- Data Science in Paris and in France
- Huawei-FSMP Mathematical Coffees
- My Past and Present Projects
- Computational Optimal Transport

TREMPLIN CARNOT Smiles

SMILES : Mathematics for Innovation

The Carnot SMILES institute aims at making the most of the power of applied mathematics in order to catalyze the innovation processes of companies in all sectors. Made up of multidisciplinary teams, it supports its industrial partners in their technological challenges by mathematical modeling and numerical simulation in: fluid and structural mechanics, acoustics, space and life sciences, big data, chemistry, electronics , plasma physics, nanosciences ...



Putting the unreasonable effectiveness of mathematics at the service of innovation within companies

Our academic partners

UPMC SORBONNE UNIVERSITÉS

université PARIS DIDEROT PARIS 7

FSMP Fondation Sciences Mathématiques de Paris

CNRS

SATT LUTECH

Why partnering with the Carnot SMILES institute ?

Security

- Scientific excellence
- Continuity of the link while running the project
- Operational and tangible output
- Transparency of the financial offer

Savings

•Flexibility of the offer in relation to the partner's need
•Saving compared to a short-term hiring on the partner' s side

Convenience

•Time saving mounting the project
•Complex combination of exceptional human resources adapted to the partner's need
•Reactivity to the partner's need
•Speed of project mounting

Why partnering with the Carnot SMILES institute ?

1. Security

- **Scientific excellence**
  - UPMC is the 5th university worldwide in mathematics (Shanghai ranking)
  - The Foundation Sciences Mathématiques de Paris, which collaborates with the institute, supplies a network of mathematicians among the most renowned and dense worldwide
- **Continuity of the link while running the project**
  - The Carnot engineer* helps the researcher carrying out the research projet
  - The managing director of the Carnot institute and the Carnot engineer* are the main representatives of the partner, the researcher being subjected to his agenda
- **Operational and tangible output**
  - The Carnot engineer* ensures the background work that complements the researcher's advanced expertise: mathematical model coding, algorithm development, report writing
- **Transparency of the financial offer**
  - A cost estimation is sent to the partner : daily cost of researcher / Carnot engineer*, differenciation between salary with taxes / environment of researchers / general expenses, so that the partner better understands the offer

* Carnot engineer : a PhD in applied mathematics, hired by the Carnot SMILES institute

Why partnering with the Carnot SMILES institute ?

2. Savings

- **Flexibility of the offer in relation to the partner's need**
  - Invoicing of the researchers' expertise upon a daily basis, according to the need predetermined together
  - Investment in breakthrough expertise, optimally used
- **Saving compared to a short-term hiring on the partner's side**
  - Time saving compared to intern recruitment (identification of the candidate, reliability of his/her intellectual value)
  - No risky hiring on the partner's side if the research need weakens or disappears

* Carnot engineer : a PhD in applied mathematics, hired by the Carnot SMILES institute

Why partnering with the Carnot SMILES institute ?

3. Convenience

- **Time saving mounting the project**
  - The managing director of the Carnot institute meets with the industrial partner in order to highlight research themes
  - Then, he identifies researchers who can answer the scientific challenges highlighted
  - Finally, he brings together the relevant experts identified from each party so that they can build a research project together
  - This way, the research experts' time is made profitable by bringing together relevant people talking about topics of common interests
- **Complex combination of exceptional human resources** adapted to the partner's need
- **Reactivity to the partner's need**
  - Immediate availability of the Carnot engineers*
  - The partner receives a cost estimate less than 3 weeks after the first meeting
  - Contractualization less than 2 months after acceptance of the cost estimate
- **Speed of project mounting**
  - Support of the managing director of the Carnot institute in order to speed up the legal process on the university side

* Carnot engineer : a PhD in applied mathematics, hired by the Carnot SMILES institute

They trust us…

ArcelorMittal

AREVA
l'avenir pour énergie

DE LA RECHERCHE À L'INDUSTRIE
cea

eDF

MICHELIN

PSA
GROUPE

SANOFI

SAFRAN
AEROSPACE · DEFENCE · SECURITY

TOTAL

URGO

…so what about you?

Overview

- FSMP
- Carnot SMILE
- **Data Science in Paris and in France**
- Huawei-FSMP Mathematical Coffees
- My Past and Present Projects
- Computational Optimal Transport

Home SEMINAR JOBS CHALLENGE DATA COMING

Data @ ENS

CFM-ENS Data Science Chair

ENS ÉCOLE NORMALE SUPÉRIEURE

PSL RESEARCH UNIVERSITY PARIS

https://data-ens.github.io

CFM INSIGHT.DATA.CLARITY.

Data Science Colloquium

of the ENS

Laplace Junior Professor Chair in Data Science

Postdoctoral Program

Challenge data

reminiz — Cluster actor faces from TV show

R — Prédire le score esthétique d'un portrait

PriceMinister Rakuten group — Prédiction de l'intérêt des avis utilisateurs

Dassault Systèmes — Filtrage collaboratif au sein de la 3DEXPERIENCE Platform

RYTHM — Prédiction de l'âge d'un sujet à partir de son activité cérébrale

Société Générale — Prédire la tendance de la production de pétrole brut

CFM INSIGHT.DATA.CLARITY. — Prediction of trading activity within the order book

VEOLIA — Prévision du risque de casse des canalisations d'un réseau

EDF — Prédire les clients qui ont réalisé des économies d'énergie

plume LABS — Prédire la qualité de l'air à l'échelle de la rue

https://imaging-in-paris.github.io

Organizers

Andres Almansa (CNRS et Telecom-Paris)
Julie Delon (Paris 5)
Agnès Desolneux (CNRS et ENS Cachan)
Jalal Fadili (ENSICAEN)
Bruno Galerne (Paris 5)
Yann Gousseau (Telecom Paris)
Gabriel Peyré (CNRS et ENS)

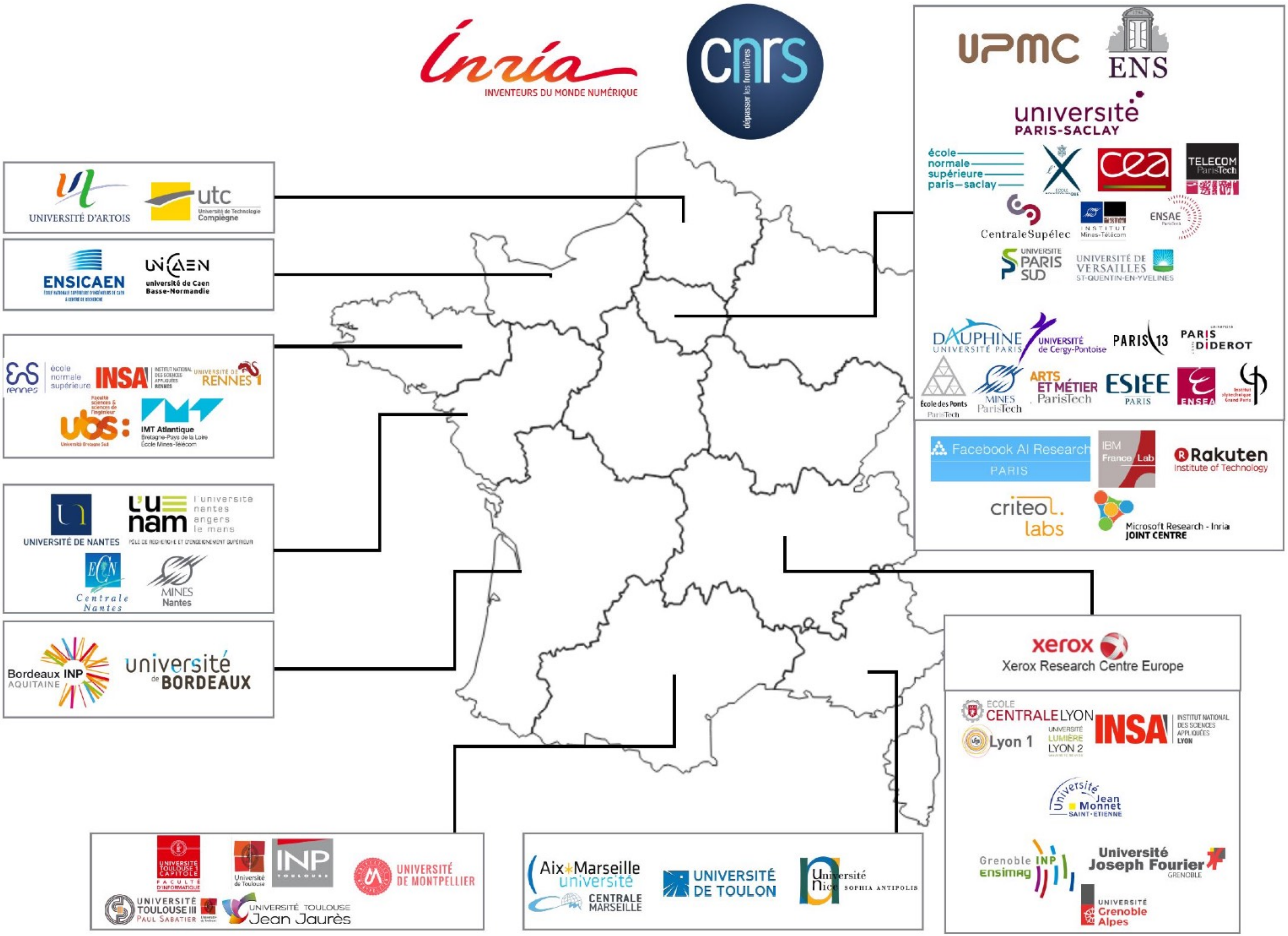
GdR MIA

SMAI SIGMA

ihp Institut Henri Poincaré

http://gdr-mia.math.cnrs.fr

Academia



Inria
INVENTEURS DU MONDE NUMÉRIQUE
cnrs
dépasser les frontières
UPMC
ENS
université PARIS-SACLAY
école normale supérieure paris—saclay
cea
TELECOM ParisTech
CentraleSupélec
ENSAE
UNIVERSITÉ PARIS SUD
UNIVERSITÉ DE VERSAILLES ST-QUENTIN-EN-YVELINES
DAUPHINE UNIVERSITÉ PARIS
UNIVERSITÉ de Cergy-Pontoise
PARIS 13
PARIS DIDEROT
École des Ponts ParisTech
MINES ParisTech
ARTS ET MÉTIER ParisTech
ESIEE PARIS
ENSEA
Facebook AI Research PARIS
IBM France Lab
Rakuten Institute of Technology
criteo labs
Microsoft Research - Inria JOINT CENTRE
UNIVERSITÉ D'ARTOIS
utc Université de Technologie Compiègne
ENSICAEN
UNICAEN université de Caen Basse-Normandie
ENS rennes école normale supérieure
INSA RENNES
UNIVERSITÉ DE RENNES 1
ubs
IMT Atlantique Bretagne-Pays de la Loire École Mines-Télécom
UNIVERSITÉ DE NANTES
L'UNAM l'université nantes angers le mans
ECN Centrale Nantes
MINES Nantes
Bordeaux INP AQUITAINE
université de BORDEAUX
xerox
Xerox Research Centre Europe
ECOLE CENTRALE LYON
INSA LYON
Lyon 1
UNIVERSITÉ LUMIÈRE LYON 2
Université Jean Monnet SAINT-ETIENNE
Grenoble INP Ensimag
Université Joseph Fourier GRENOBLE
UNIVERSITÉ Grenoble Alpes
UNIVERSITÉ TOULOUSE 1 CAPITOLE
INP TOULOUSE
UNIVERSITÉ DE MONTPELLIER
UNIVERSITÉ TOULOUSE III PAUL SABATIER
UNIVERSITÉ TOULOUSE Jean Jaurès
Aix*Marseille université
CENTRALE MARSEILLE
UNIVERSITÉ DE TOULON
Université Nice SOPHIA ANTIPOLIS

Education

Fields medal:
2nd country with ENS #1 worldwide institution

14

13 France: #2 country

9

ENS #1 institution worldwide with 11 medals

Universities:
5 in the top 30 math universities (1 in top 5)*

UPMC, Université Paris-Saclay, ENS, Aix-Marseille Université, Dauphine Université Paris

Engineering schools:
200+ schools graduating 38,000 new engineers per year with a very advanced math knowledge. Many schools have dedicated Machine Learning / Data Scientist programs.

Telecom ParisTech, ENS Cachan, ENS, Centrale, l'X École Polytechnique, Grenoble INP Ensimag, ENSTA ParisTech Université Paris-Saclay, Mines ParisTech, École des Ponts ParisTech, Efrei Paris Sud, EPITA, ENSAE ParisTech

MVA Mathématiques / Vision / Apprentissage

École normale supérieure Paris-Saclay, CentraleSupélec, École des Ponts, ENSAE ParisTech, ENSTA, Mines Paris ParisTech, l'X École Polytechnique Université Paris-Saclay, Telecom ParisTech, Université Paris Descartes, Dauphine Université Paris

Master Data Sciences

Université Paris-Saclay, Telecom ParisTech, ENSAE ParisTech, l'X École Polytechnique Université Paris-Saclay, Université Paris Sud

Communities

ROBOT LAB
TELECOM ParisTech
Agoranov

Incubators

Paris-Saclay Center for Data Science
scikit learn
TERALAB
IPOL IMAGE PROCESSING ON LINE
DATA GOUV.FR

Open source Plateforms

MACHINE LEARNING POWER TO THE DATA
Deep Learning Paris Meetup
APPRENTISSAGE PROFOND
Data @ ENS CFM-ENS Data Science Chair
Paris Ai Robotics IoT Meetup
Machine Learning Aix - Marseille
Mulhouse Machine Learning
BORDEAUX MACHINE LEARNING
Chatbots Paris
Paris Tech Talks
GdR MIA
DATA FOR GOOD X BAYES IMPACT
TDS
Paris Datageeks
BigData MONTPELLIER
Kaggle Paris Meetup
Grenoble Data Science

Meetings

Startups

Agents

Personal
jam, snips, Wiidii

Professional
Birdly, Julie Desk, Uli.ai, ottspott

Platforms / Tools
XBRAIN, clustaar., davi, craft ai, Botfuel, Smartly.ai, Recast.ai

Robotics (Hardware + Software)
Robocortex, Immersive Robotics, Intuitive Robots, Blue Frog, Axyn, Aldebaran, Stanley Robotics, Event Bots, Hoomano, Stimul'activ, MWOO, Octopus Robots, GOBIO Robot, Partnering Robotics, naïo

Enterprise

Security / Fraud
QOSMOS, Shift Technology, CybelAngel, ITrust, Cypheme, Alkheros

Marketing / Sales
tinyclues, CustomerMatrix, BEYABLE, GEOMAIL, Search xpr, tastehit, mediego

Customer support
gorgias, XBRAIN, CallDesk, SENDERS, davi, viavoo, DoYouDreamUp, dictaNOVA

Business + Predictive Intelligence
linkfluence, PredicSis, ZELROS, clustaar., DATAPRED, antidot, IDMOG

HR
Clustree, riminder

Platforms

Hardware / Processing
Chronocam, LightOn, Kalray, Cortex

Big Data / Machine Learning
dataiku, SINEQUA, craft ai, antidot, tellmeplus, exensa, DATAPRED, Invenis

Natural Language
proxem, syJLabs, yseop

Vision / Image
Chronocam, Angus.ai, Smartmeup, deepomatic, heuritech

Audio
niland, BatVoice Technologies, ava, OtoSense

Industries

AdTech
DataBerries, adomik

Insurance
Shift Technology, alan, davi, DreamQuark

Healthcare / Pharma
Cardiologs, Khresterion, BIOSERENITY, HONESTICA, DreamQuark, MENSIA Technologies, :implicity, PIXYL, DEXSTR, Qynapse

Retail
ALKEMICS, verteego, Moodstocks, madd.ie, tinyclues, Neosensys, deepomatic, placemeter, AntVoice, Occi, tastehit, Search xpr

Legal
Doctrine.fr

Manufacturing / Logistics
BALYO, Vekia, Cortex, energiency

Transportation
nauya, karos, EasyMile, vulog, AutoKAB

City + Infrastructures
plume, DCbrain, OtoSense, placemeter

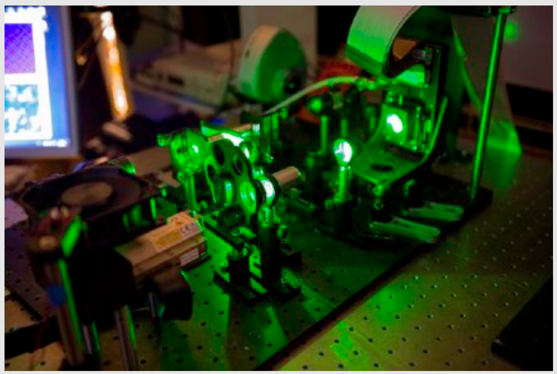
http://franceisai.com/startups

Biased Sub-sampling

LightOn

www.lighton.io

Optical Computing for Machine Learning



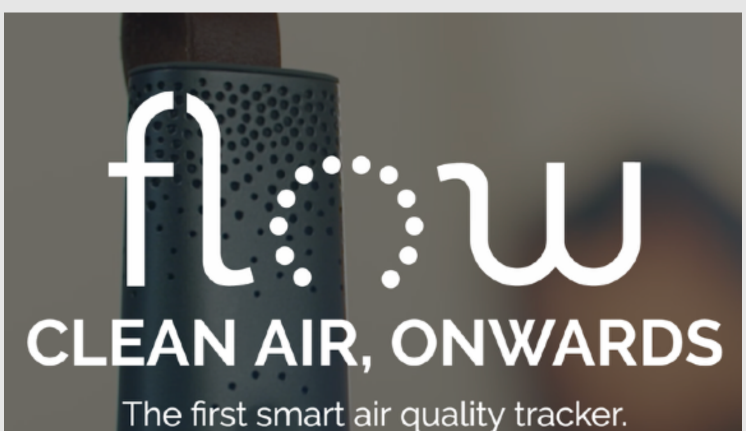
REGAIND

www.regaind.io

Extract game-changing insights from images,
increase your activity and reduce your costs

plume LABS

www.plumelabs.com



yelloan

www.yelloan.com

Data science and social network
for credit assessment

Overview

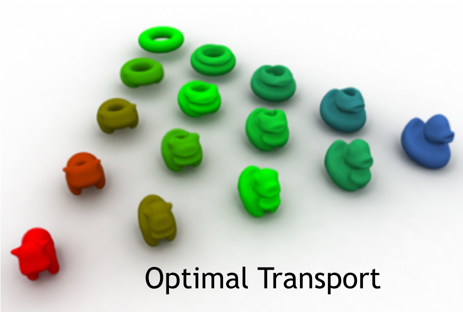
- FSMP
- Carnot SMILE
- Data Science in Paris and in France
- **Huawei-FSMP Mathematical Coffees**
- My Past and Present Projects
- Computational Optimal Transport

Mathematical Coffees

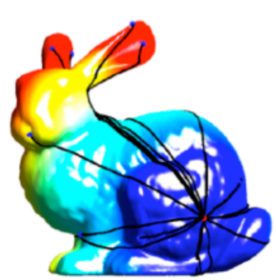
Huawei-FSMP joint seminars

https://mathematical-coffees.github.io

Organized by: Mérouane Debbah & Gabriel Peyré

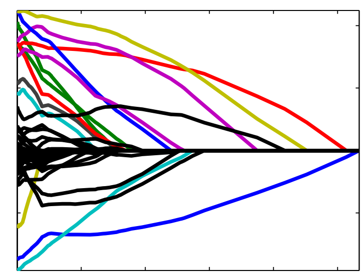


Optimal Transport

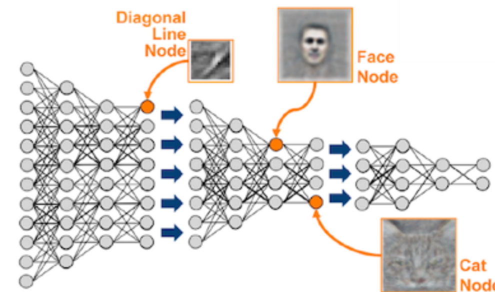


Geodesics

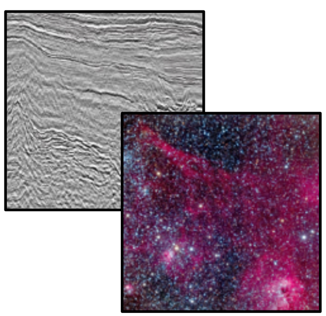
Meshes



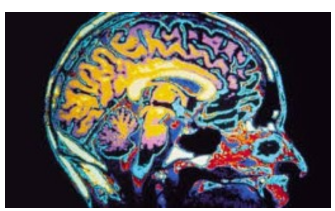
Optimization



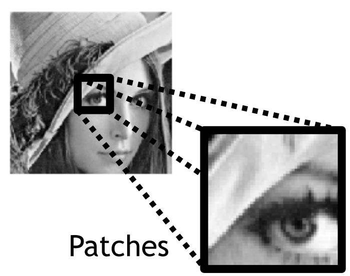
Deep Learning



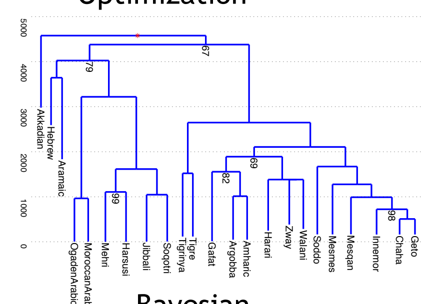
Sparsity



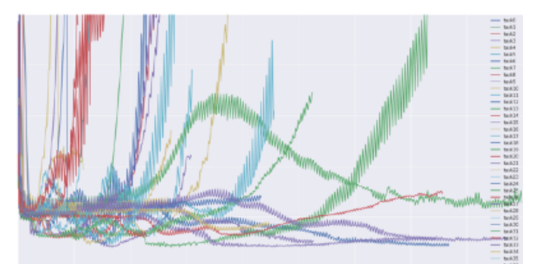
Neuro-imaging



Patches



Bayesian



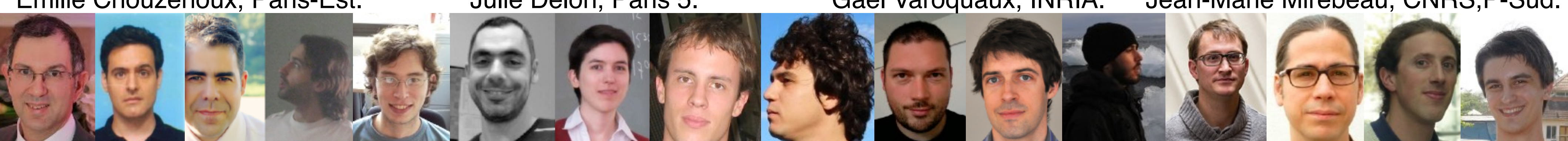
Parallel/Stochastic

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.
Julien Tierny, CNRS and P6.
Robin Ryder, Paris-Dauphine.
Gael Varoquaux, INRIA.

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

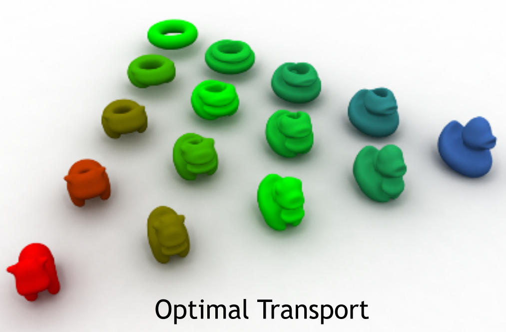


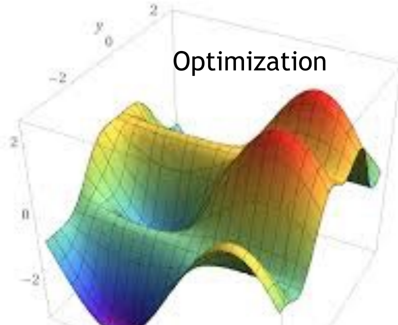
Mathematical Coffees

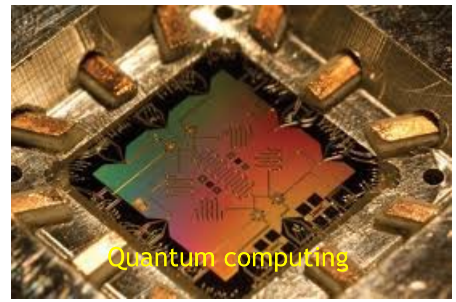
Huawei-FSMP joint seminars

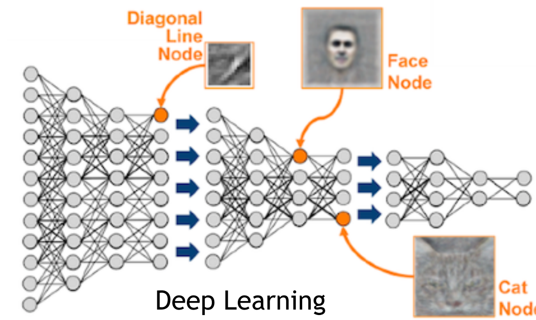
https://mathematical-coffees.github.io

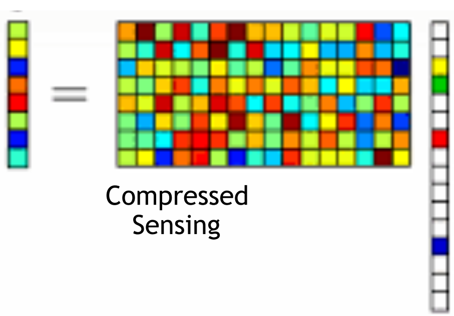
Organized by: Mérouane Debbah & Gabriel Peyré

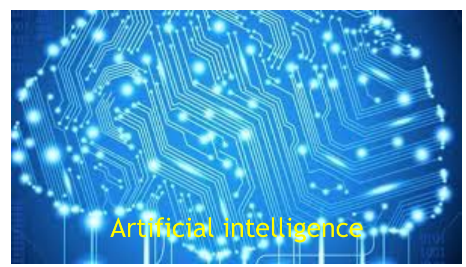


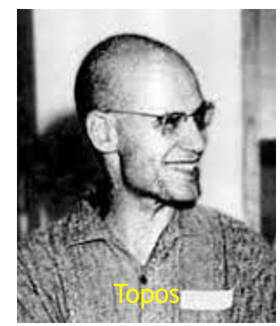












Yves Achdou, Paris 6
Daniel Bennequin, Paris 7
Marco Cuturi, ENSAE
Jalal Fadili, ENSICaen

Alexandre Gramfort, INRIA
Olivier Grisel (INRIA)
Olivier Guéant, Paris 1
Iordanis Kerenidis, CNRS and Paris 7
Guillaume Lecué, CNRS and ENSAE

Frédéric Magniez, CNRS and Paris 7
Edouard Oyallon, CentraleSupelec
Gabriel Peyré, CNRS and ENS
Joris Van den Bossche (INRIA)

Model Fitting in Data Sciences

$$\min_{\theta} \mathcal{E}(\theta) \overset{\text{def.}}{=} L(f(x, \theta), y)$$

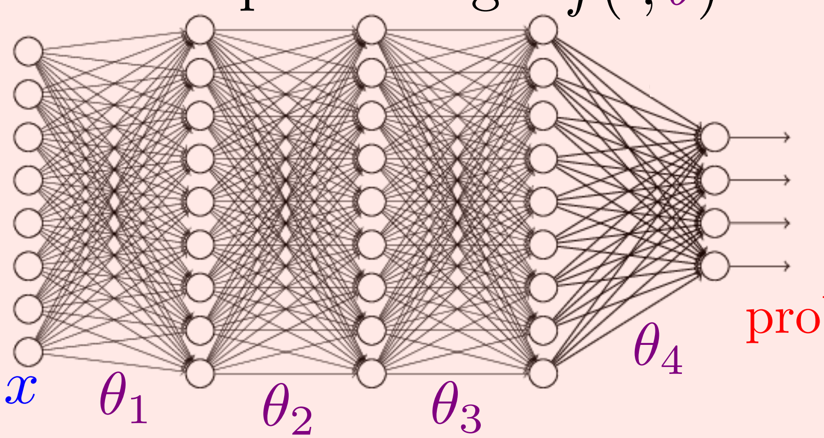
Loss Model Input Parameter Output

Model Fitting in Data Sciences

$$\min_{\theta} \mathcal{E}(\theta) \overset{\text{def.}}{=} L(f(x,\theta),y)$$

Loss Model Input Parameter Output

Deep-learning: $f(\cdot,\theta)$

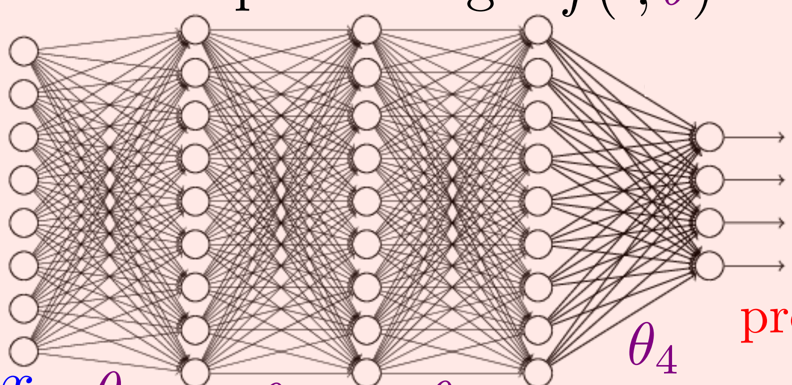


Model Fitting in Data Sciences

$$\min_{\theta} \mathcal{E}(\theta) \overset{\text{def.}}{=} L(f(x,\theta),y)$$

Loss Model Input Parameter Output

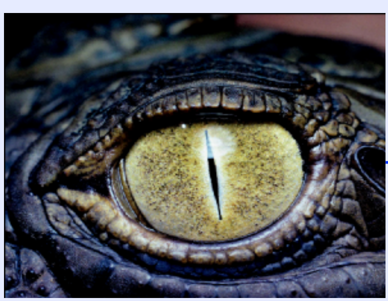
Deep-learning: $f(\cdot,\theta)$



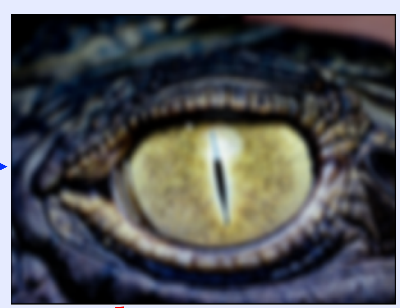
x θ_1 θ_2 θ_3 θ_4

y class probabilities

Super-resolution:



$f(x,\cdot)$ degradation



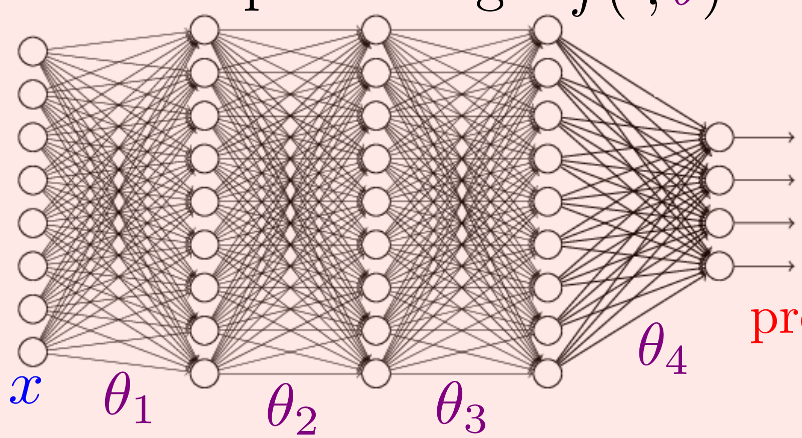
θ unknown image y observation

Model Fitting in Data Sciences

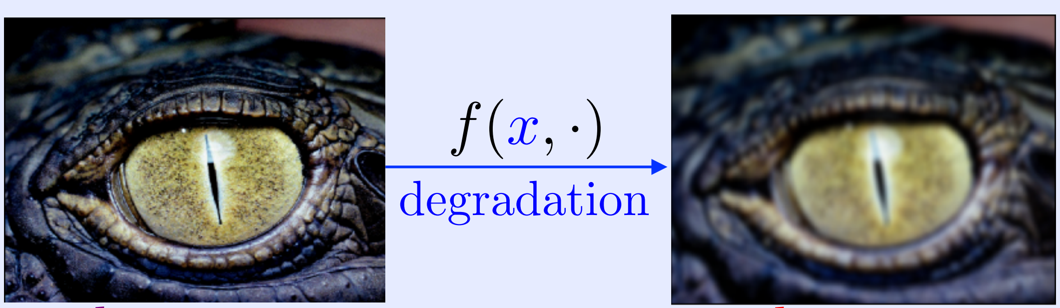
$$\min_{\theta} \mathcal{E}(\theta) \overset{\text{def.}}{=} L(f(x,\theta),y)$$

Loss Model Input Parameter Output

Deep-learning: $f(\cdot,\theta)$

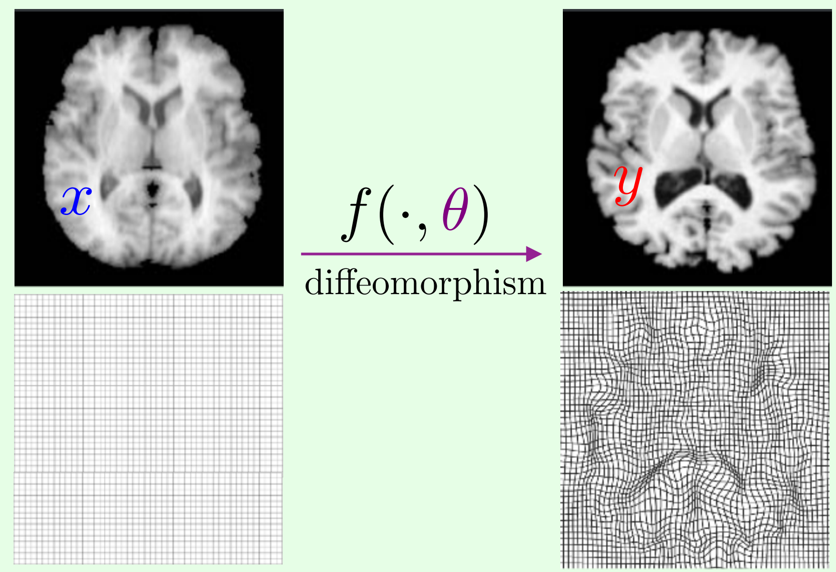


Super-resolution:



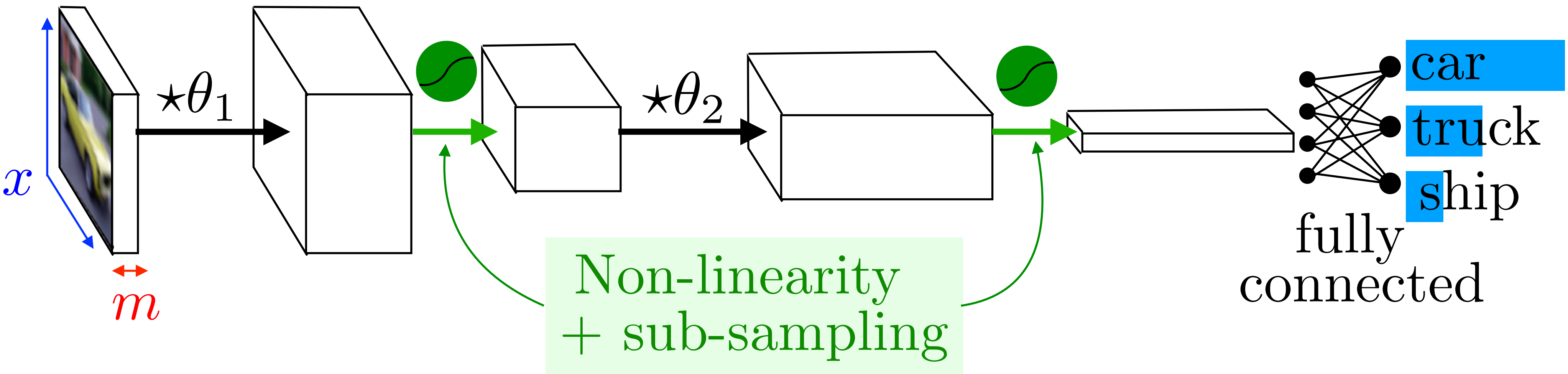
θ unknown image y observation

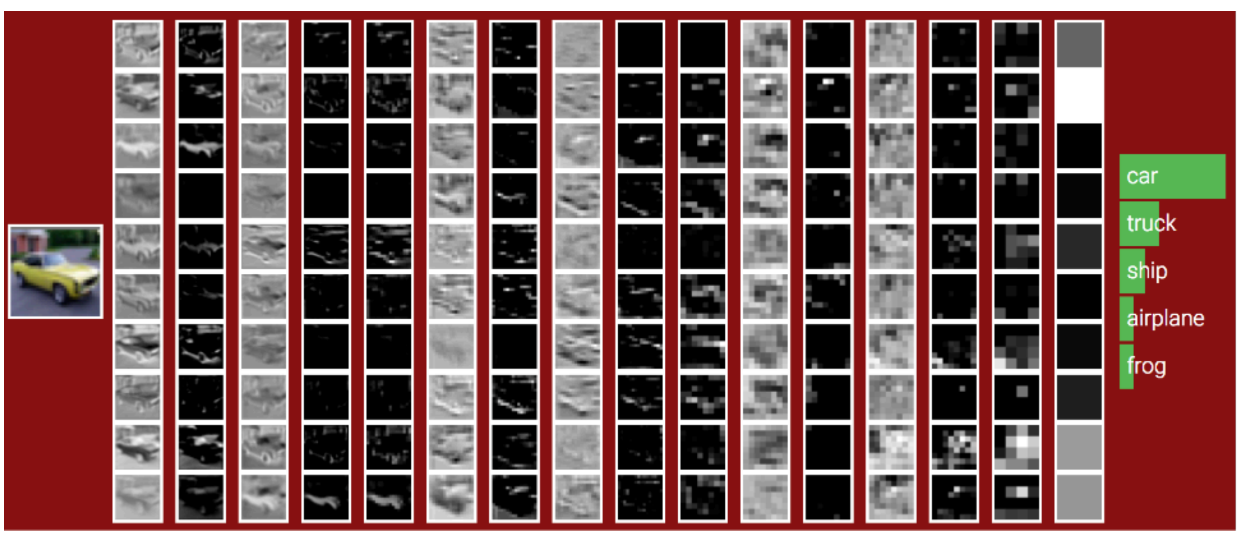
Medical imaging registration:



Multi-canal convolution: $(f \star \theta)_m(x) \stackrel{\text{def.}}{=} \sum_{\ell} \sum_{y+z=x} f_\ell(y)\theta_{m,\ell}(z)$

canal position



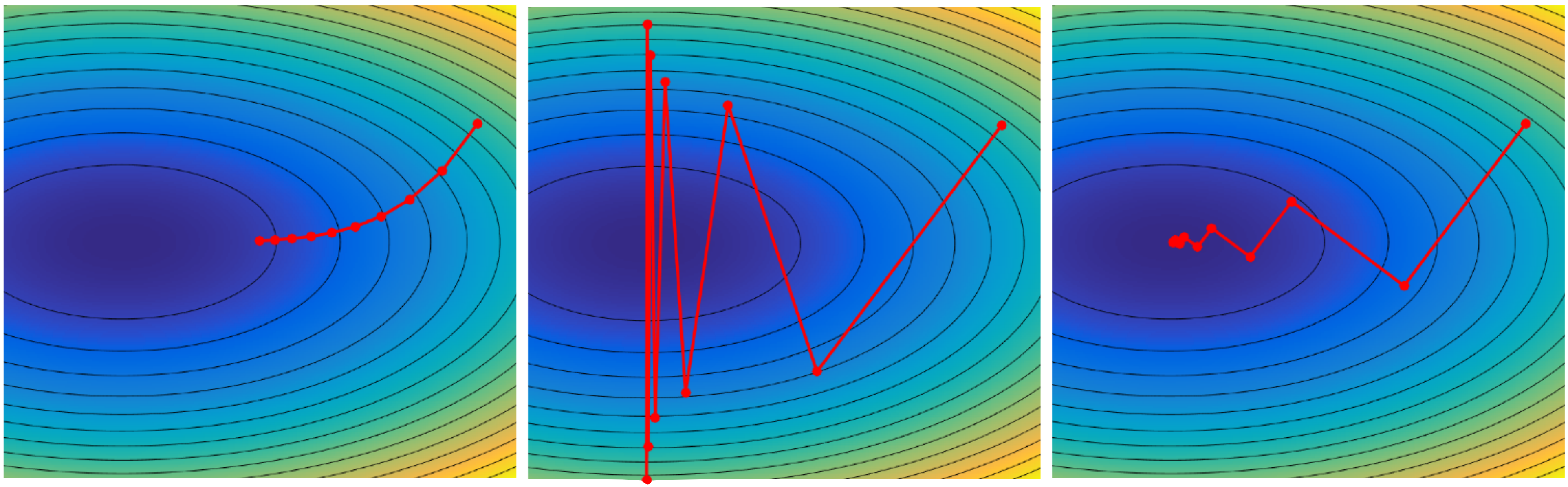


http://vision.stanford.edu/teaching/cs231n/

Gradient-based Methods

$$\min_{\theta} \mathcal{E}(\theta) \overset{\text{def.}}{=} L(f(x,\theta), y)$$

$$\text{Gradient descent:} \quad \theta_{\ell+1} = \theta_\ell - \tau_\ell \nabla \mathcal{E}(\theta_\ell)$$

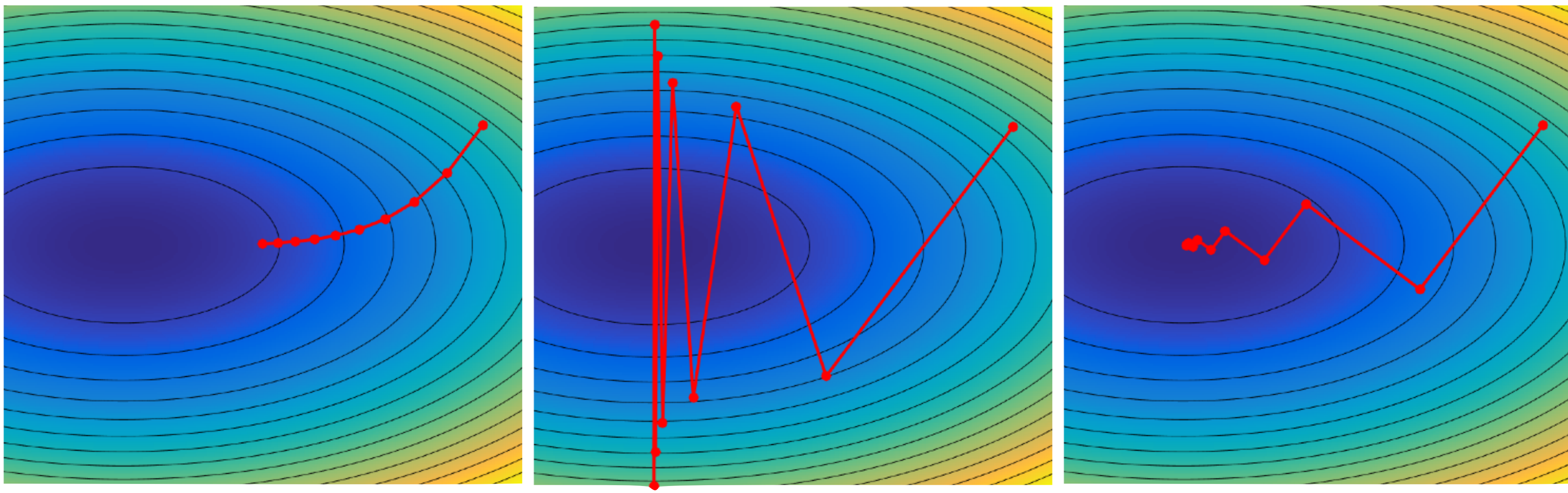


Small τ_ℓ | Large τ_ℓ | Optimal $\tau_\ell = \tau_\ell^\star$

Gradient-based Methods

$$\min_{\theta} \mathcal{E}(\theta) \overset{\text{def.}}{=} L(f(x,\theta),y)$$

$$\text{Gradient descent:} \quad \theta_{\ell+1} = \theta_\ell - \tau_\ell \nabla \mathcal{E}(\theta_\ell)$$



Small τ_ℓ | Large τ_ℓ | Optimal $\tau_\ell = \tau_\ell^\star$

Many generalization:

- Nesterov / heavy-ball
- (quasi)-Newton
- Stochastic / incremental methods
- Proximal splitting (non-smooth $\mathcal{E}$)
- $\ldots$

The Complexity of Gradient Computation

Setup: $\mathcal{E} : \mathbb{R}^n \to \mathbb{R}$ computable in K operations.

```
def ForwardNN(A,b,Z):
    X = []
    X.append(Z)
    for r in arange(0,R):
        X.append( rhoF( A[r].dot(X[r]) + tile(b[r],[1,Z.shape[1]]) ) )
    return X
```

Hypothesis: elementary operations $(a \times b, \log(a), \sqrt{a} \ldots)$ and their derivatives cost $O(1)$.

The Complexity of Gradient Computation

Setup: $\mathcal{E} : \mathbb{R}^n \to \mathbb{R}$ computable in K operations.

```
def ForwardNN(A,b,Z):
    X = []
    X.append(Z)
    for r in arange(0,R):
        X.append( rhoF( A[r].dot(X[r]) + tile(b[r],[1,Z.shape[1]]) ) )
    return X
```

Hypothesis: elementary operations $(a \times b, \log(a), \sqrt{a} \ldots)$ and their derivatives cost $O(1)$.

Question: What is the complexity of computing $\nabla \mathcal{E} : \mathbb{R}^n \to \mathbb{R}^n$?

The Complexity of Gradient Computation

Setup: $\mathcal{E} : \mathbb{R}^n \to \mathbb{R}$ computable in K operations.

```
def ForwardNN(A,b,Z):
    X = []
    X.append(Z)
    for r in arange(0,R):
        X.append( rhoF( A[r].dot(X[r]) + tile(b[r],[1,Z.shape[1]]) ) )
    return X
```

Hypothesis: elementary operations $(a \times b, \log(a), \sqrt{a} \dots)$ and their derivatives cost $O(1)$.

Question: What is the complexity of computing $\nabla\mathcal{E} : \mathbb{R}^n \to \mathbb{R}^n$?

Finite differences: $$\nabla\mathcal{E}(\theta) \approx \frac{1}{\varepsilon}(\mathcal{E}(\theta+\varepsilon\delta_1) - \mathcal{E}(\theta), \dots \mathcal{E}(\theta+\varepsilon\delta_n) - \mathcal{E}(\theta))$$

$K(n+1)$ operations, intractable for large n.

The Complexity of Gradient Computation

Setup: $\mathcal{E} : \mathbb{R}^n \to \mathbb{R}$ computable in K operations.

```
def ForwardNN(A,b,Z):
    X = []
    X.append(Z)
    for r in arange(0,R):
        X.append( rhoF( A[r].dot(X[r]) + tile(b[r],[1,Z.shape[1]]) ) )
    return X
```

Hypothesis: elementary operations $(a \times b, \log(a), \sqrt{a} \ldots)$ and their derivatives cost $O(1)$.

Question: What is the complexity of computing $\nabla\mathcal{E} : \mathbb{R}^n \to \mathbb{R}^n$?

Finite differences: $$\nabla\mathcal{E}(\theta) \approx \frac{1}{\varepsilon}(\mathcal{E}(\theta + \varepsilon\delta_1) - \mathcal{E}(\theta), \ldots \mathcal{E}(\theta + \varepsilon\delta_n) - \mathcal{E}(\theta))$$

$K(n+1)$ operations, intractable for large n.

Theorem: there is an algorithm to compute $\nabla\mathcal{E}$ in $O(K)$ operations. [Seppo Linnainmaa, 1970]

The Complexity of Gradient Computation

Setup: $\mathcal{E} : \mathbb{R}^n \to \mathbb{R}$ computable in K operations.

```
def ForwardNN(A,b,Z):
    X = []
    X.append(Z)
    for r in arange(0,R):
        X.append( rhoF( A[r].dot(X[r]) + tile(b[r],[1,Z.shape[1]]) ) )
    return X
```

Hypothesis: elementary operations $(a \times b, \log(a), \sqrt{a} \ldots)$
and their derivatives cost $O(1)$.

Question: What is the complexity of computing $\nabla\mathcal{E} : \mathbb{R}^n \to \mathbb{R}^n$?

Finite differences: $$\nabla\mathcal{E}(\theta) \approx \frac{1}{\varepsilon}(\mathcal{E}(\theta + \varepsilon\delta_1) - \mathcal{E}(\theta), \ldots \mathcal{E}(\theta + \varepsilon\delta_n) - \mathcal{E}(\theta))$$
$K(n+1)$ operations, intractable for large n.

Theorem: there is an algorithm to compute $\nabla\mathcal{E}$ in $O(K)$ operations.
[Seppo Linnainmaa, 1970]

This algorithm is reverse mode automatic differentiation

```
def BackwardNN(A,b,X):
    gx = lossG(X[R],Y) # initialize the gradient
    for r in arange(R-1,-1,-1):
        M = rhoG( A[r].dot(X[r]) + tile(b[r],[1,n]) ) * gx
        gx = A[r].transpose().dot(M)
        gA[r] =  M.dot(X[r].transpose())
        gb[r] =  MakeCol(M.sum(axis=1))
    return [gA,gb]
```

Softwares

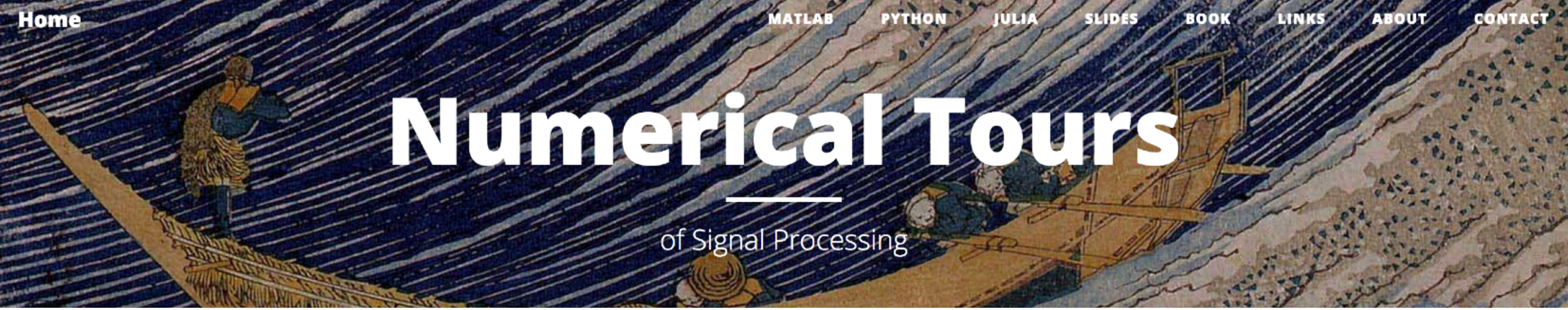
PYTORCH

TensorFlow

Chainer

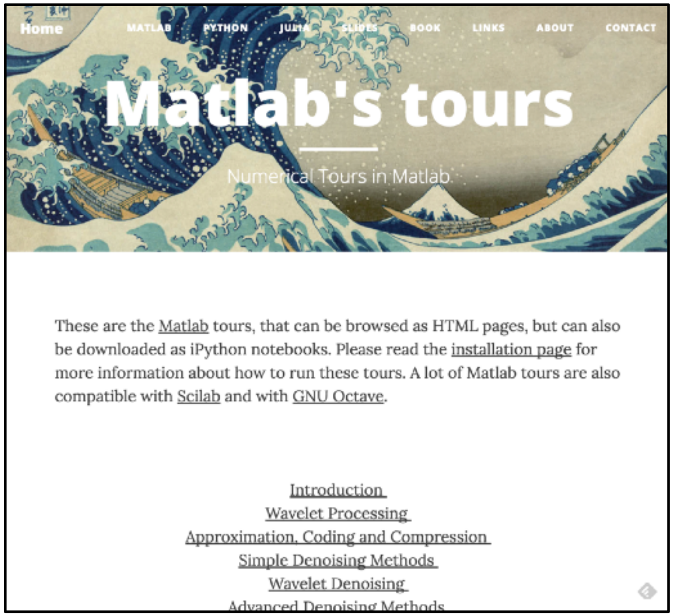
Overview

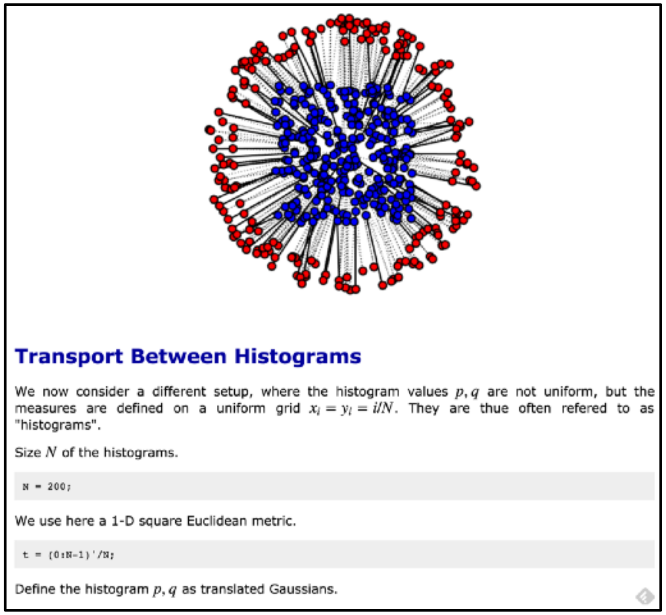
- FSMP
- Carnot SMILE
- Data Science in Paris and in France
- Huawei-FSMP Mathematical Coffees
- **My Past and Present Projects**
- Computational Optimal Transport

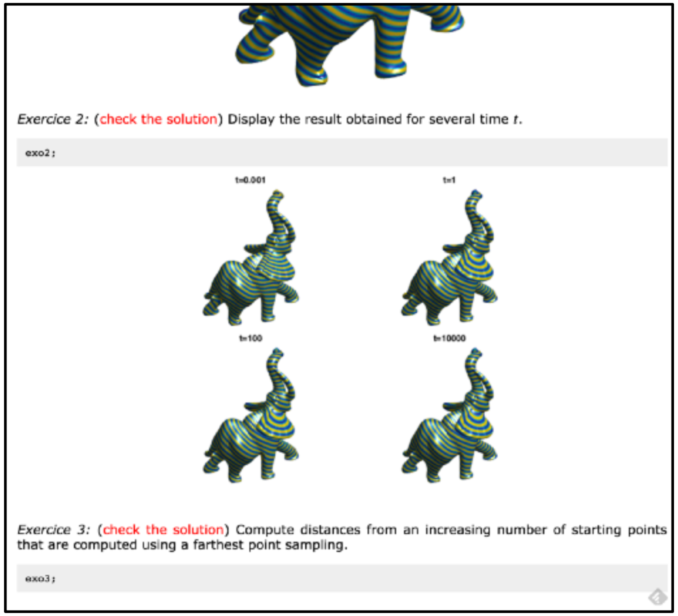


www.numerical-tours.com

Introduction	Optimal Transport
Wavelet Processing	Sparse Representations
Approximation, Coding, Compression	Inverse Problems
Simple Denoising Methods	Partial Differential Equations
Wavelet Denoising	Segmentation and Curves
Advanced Denoising Methods	Geodesic Processing
Audio Processing	Shapes
Higher Dimensional Processing	Mesh Processing
Computer Graphics	Mesh Parameterization
Optimization	Multiscale Mesh Processing



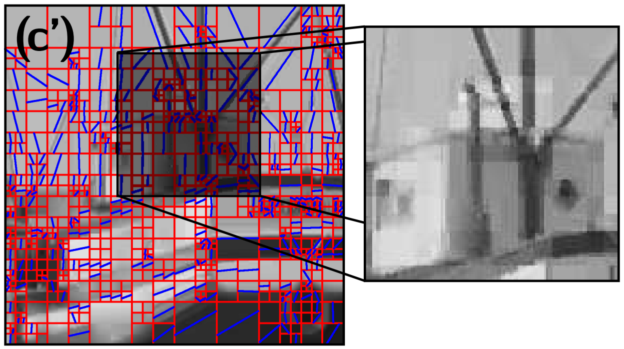




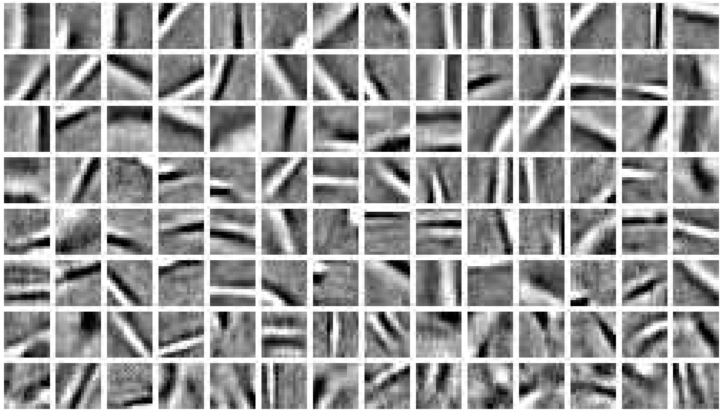
NatImages

AGENCE NATIONALE DE LA RECHERCHE ANR

2009-2011

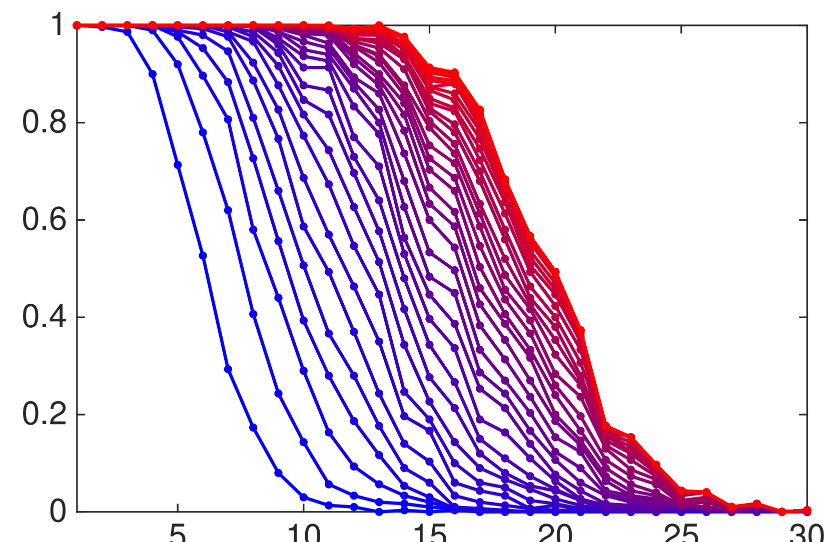


Wavelets/bandlets



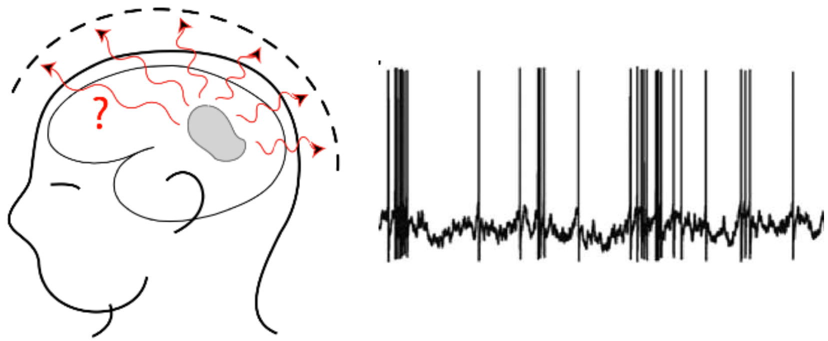
Dictionary learning
Texture synthesis

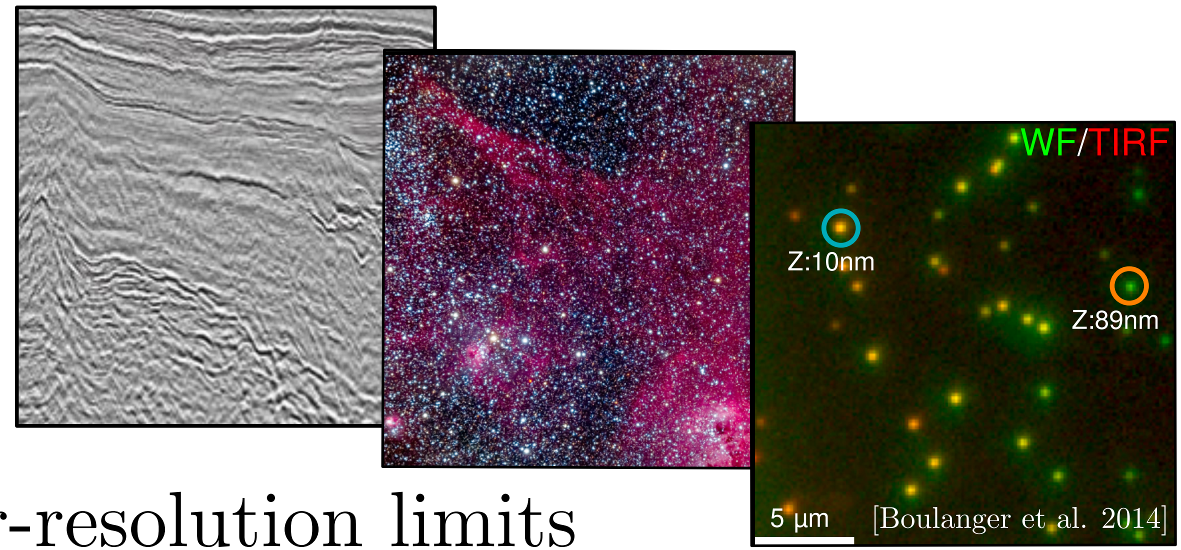
Meshes and geodesics



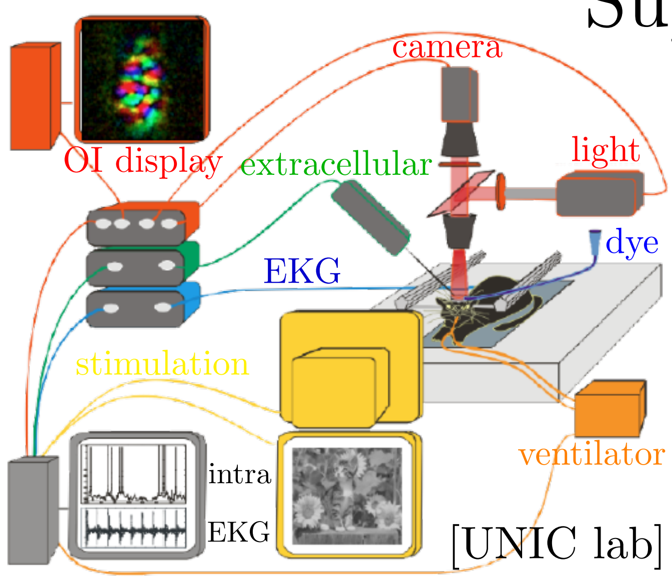
Compressed sensing

2012-2016

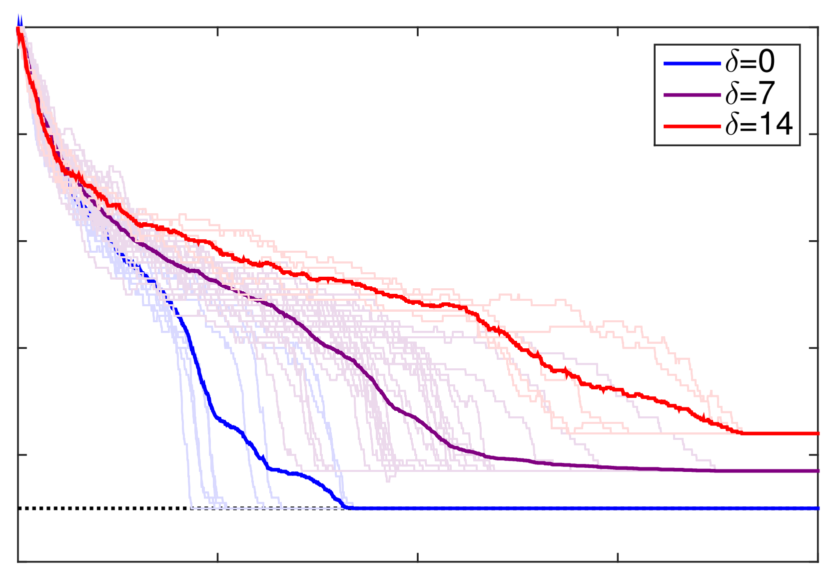




Super-resolution limits



Neurosciences of vision

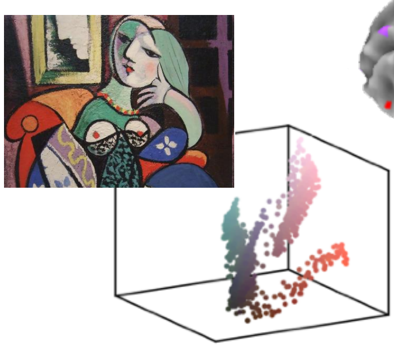
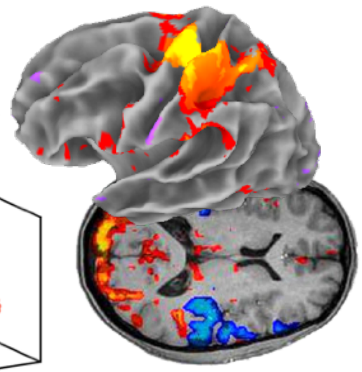
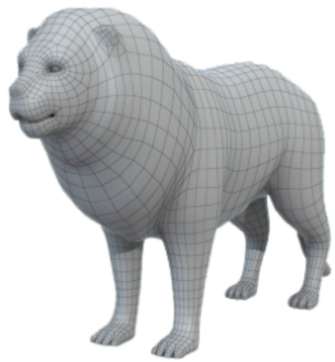


Convex optimization

noria

erc 2017-2021

Data modelling using histograms and distributions.


Optimal Transport theory.

Monge

Kantorovic

Nobel'7

Dantzig

Wasserstein

Brenier

Otto

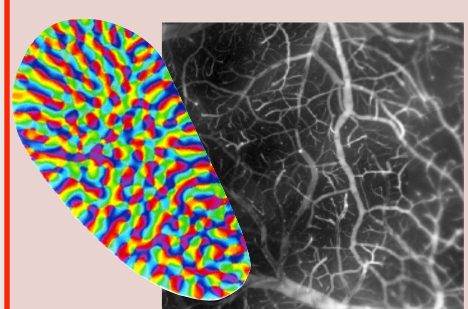
McCann

Fields'10

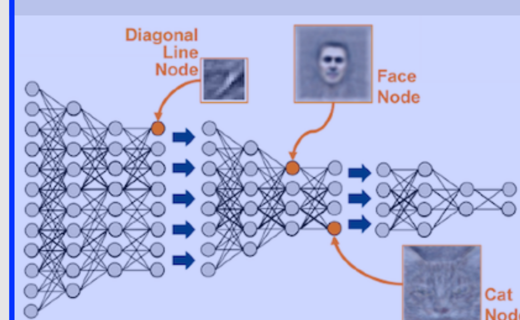
Villani

Applications.

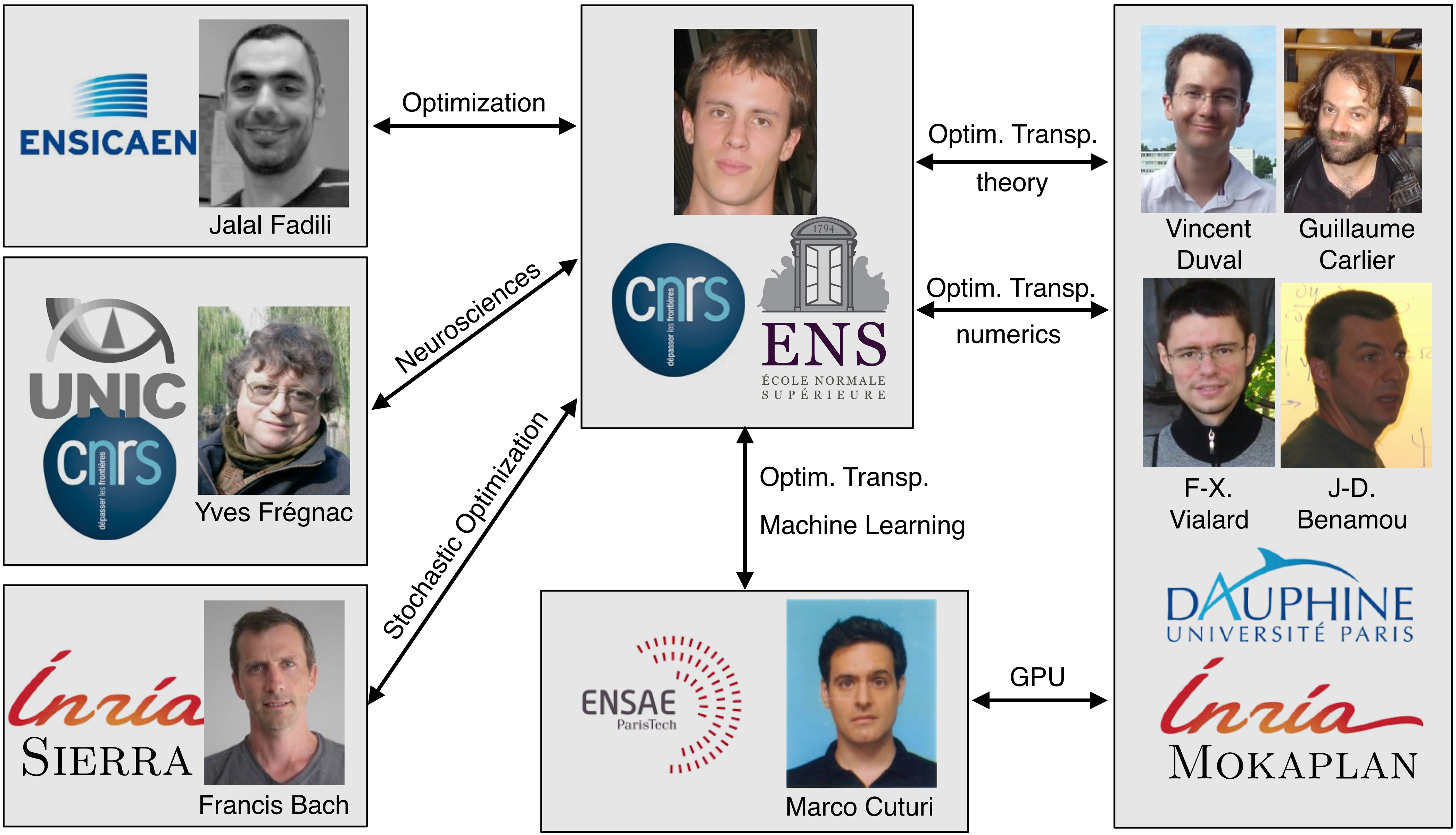
Neurosciences



Machine learning



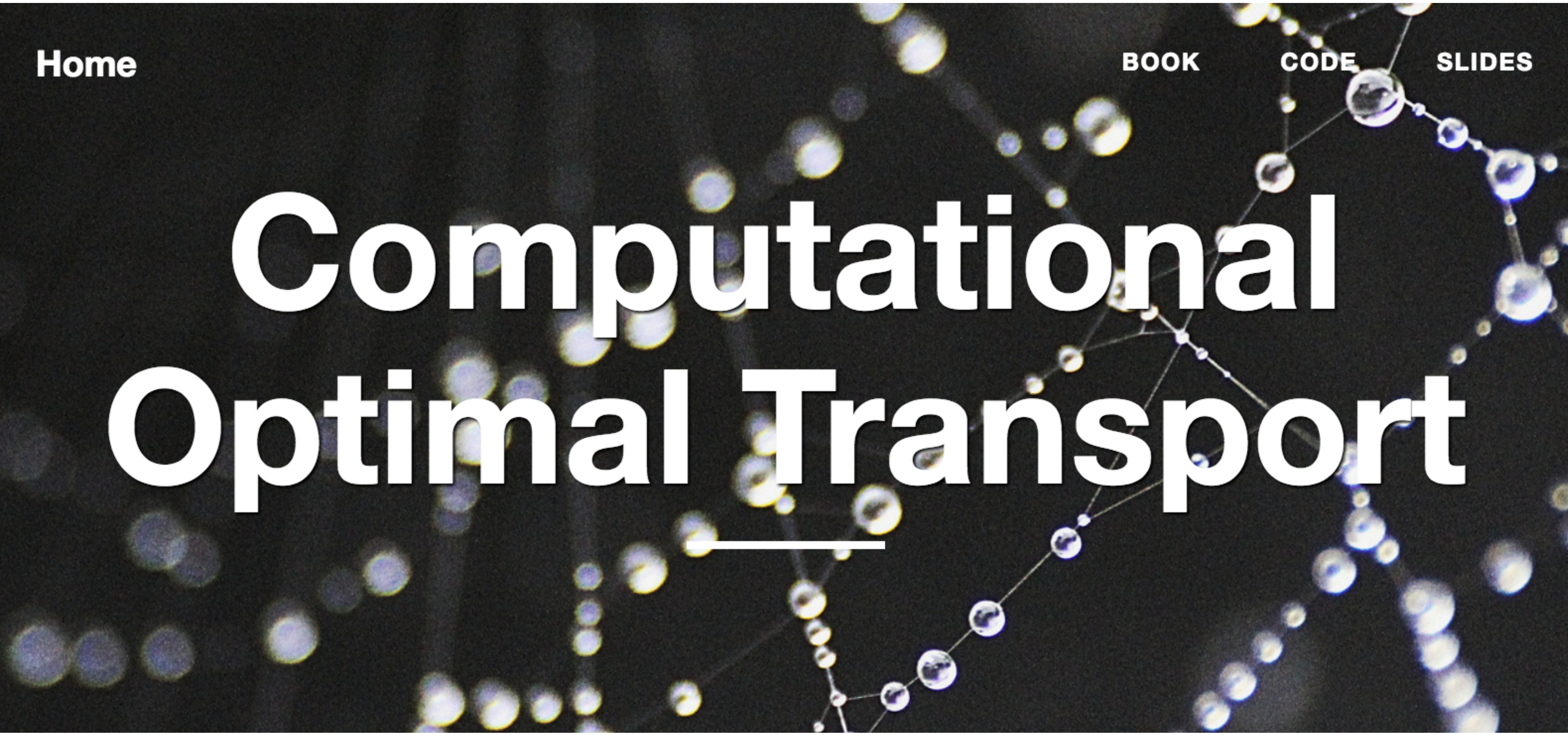
noria Ecosystem



Overview

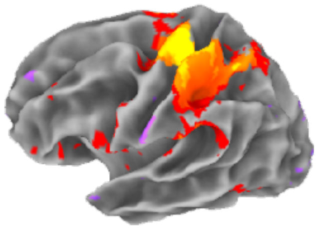
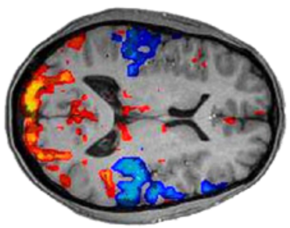
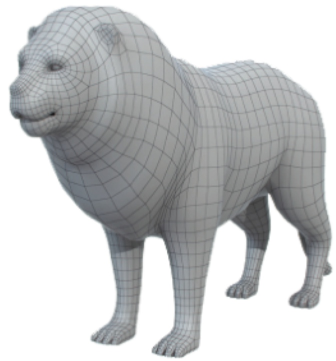
- FSMP
- Carnot SMILE
- Data Science in Paris and in France
- Huawei-FSMP Mathematical Coffees
- My Past and Present Projects
- Computational Optimal Transport

https://optimaltransport.github.io



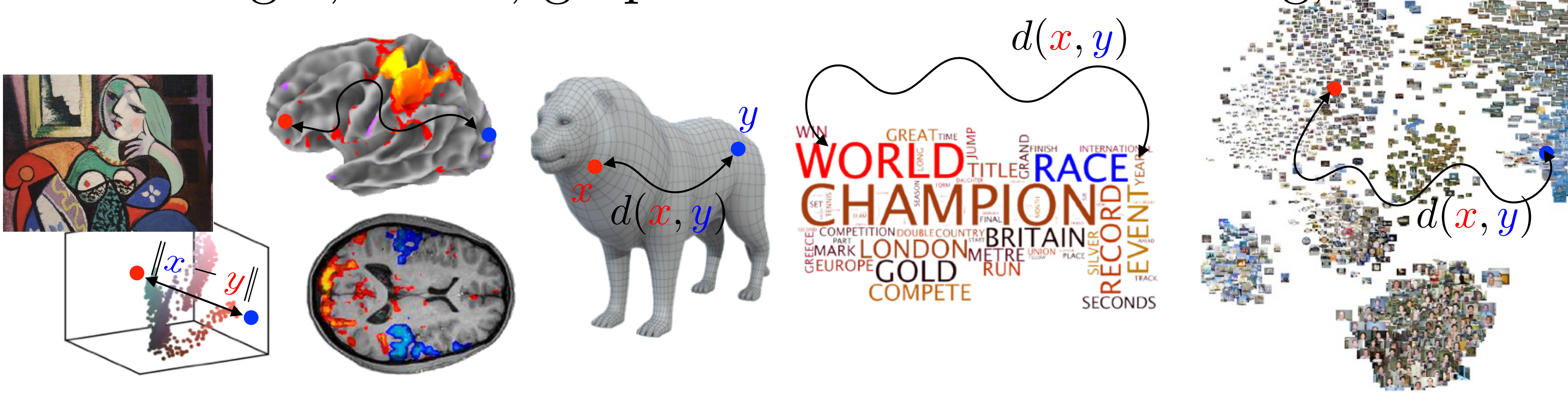
Probability Distributions in Imaging and ML

- *Probability distributions and histograms*
  $\rightarrow$ images, vision, graphics and machine learning, . . .

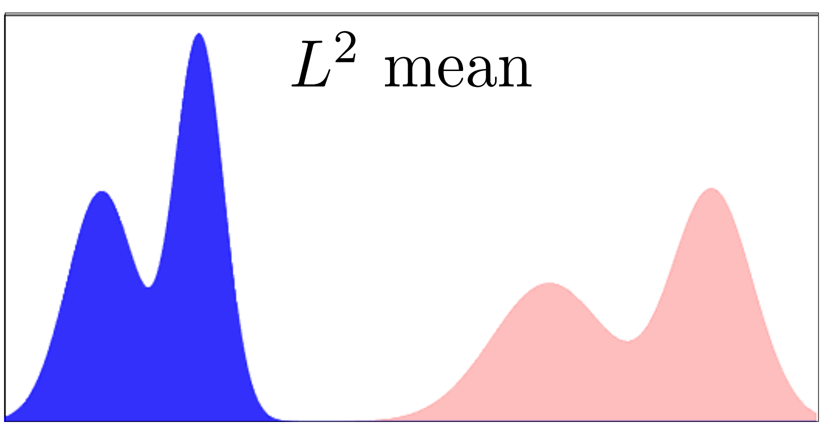





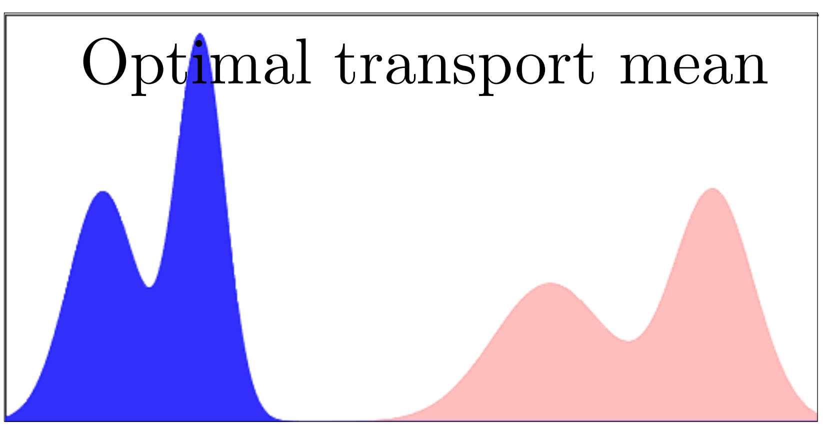

Probability Distributions in Imaging and ML

- *Probability distributions and histograms*
  $\rightarrow$ images, vision, graphics and machine learning, . . .



- *Optimal transport*
  $\rightarrow$ takes into account a metric d.





Gaspard Monge (1746-1818)

[1776]

MÉMOIRE

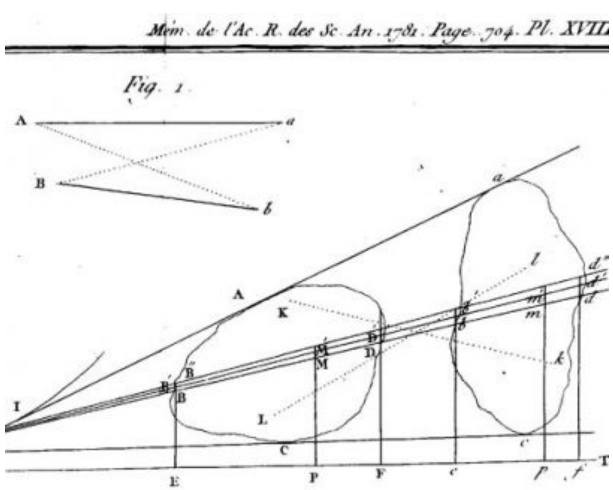
SUR LA

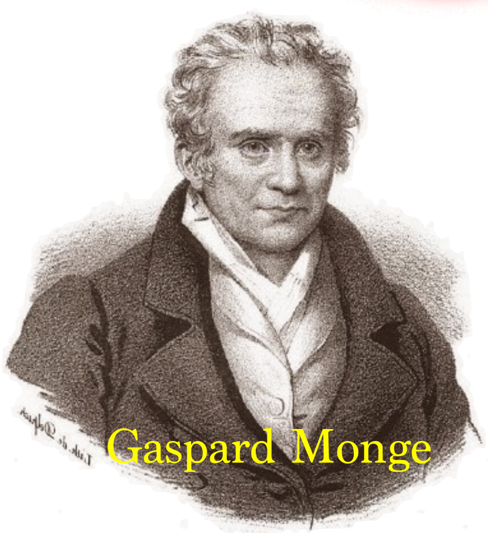
THÉORIE DES DÉBLAIS ET DES REMBLAIS.

Par M. MONGE.

LORSQU'ON doit tranſporter des terres d'un lieu dans un autre, on a coutume de donner le nom de *Déblai* au volume des terres que l'on doit tranſporter, & le nom de *Remblai* à l'eſpace qu'elles doivent occuper après le tranſport.

Le prix du tranſport d'une molécule étant, toutes choſes d'ailleurs égales, proportionnel à ſon poids & à l'eſpace qu'on lui fait parcourir, & par conſéquent le prix du tranſport total devant être proportionnel à la ſomme des produits des molécules multipliées chacune par l'eſpace parcouru, il s'enſuit que le déblai & le remblai étant donnés de figure & de poſition, il n'eſt pas indifférent que telle molécule du déblai ſoit tranſportée dans tel ou tel autre endroit du remblai, mais qu'il y a une certaine diſtribution à faire des molécules du premier dans le ſecond, d'après laquelle la ſomme de ces produits ſera la moindre poſſible, & le prix du tranſport total ſera un *minimum*.





Gaspard Monge (1746-1818)

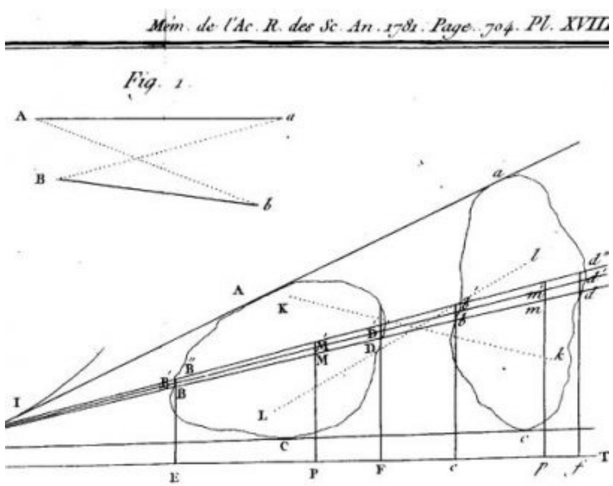
[1776]

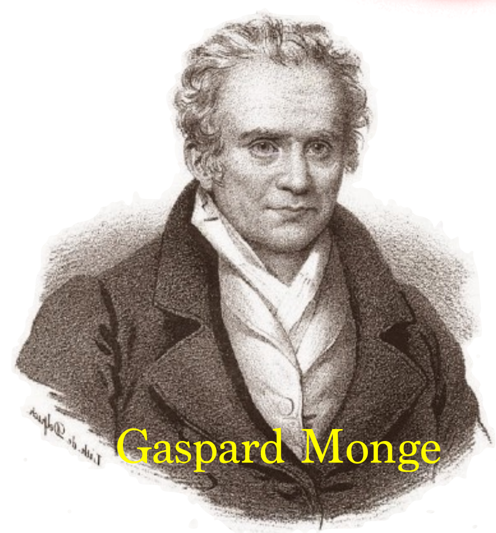
MÉMOIRE
SUR LA
THÉORIE DES DÉBLAIS
ET DES REMBLAIS.

Par M. MONGE.

LORSQU'ON doit tranſporter des terres d'un lieu dans un autre, on a coutume de donner le nom de *Déblai* au volume des terres que l'on doit tranſporter, & le nom de *Remblai* à l'eſpace qu'elles doivent occuper après le tranſport.

Le prix du tranſport d'une molécule étant, toutes choſes d'ailleurs égales, proportionnel à ſon poids & à l'eſpace qu'on lui fait parcourir, & par conſéquent le prix du tranſport total devant être proportionnel à la ſomme des produits des molécules multipliées chacune par l'eſpace parcouru, il s'enſuit que le déblai & le remblai étant donnés de figure & de poſition, il n'eſt pas indifférent que telle molécule du déblai ſoit tranſportée dans tel ou tel autre endroit du remblai, mais qu'il y a une certaine diſtribution à faire des molécules du premier dans le ſecond, d'après laquelle la ſomme de ces produits ſera la moindre poſſible, & le prix du tranſport total ſera un *minimum*.





Leonid Kantorovich (1912-1986)

Л. В. Канторович

О ПЕРЕМЕЩЕНИИ МАСС

Мы будем считать R метрическим компактным пространством, хотя некоторые из приведенных определений и результатов могут быть высказаны и для пространств более общего вида.

Пусть $\Phi(e)$ распределение масс, т.е. функция совокупности: 1) определенная для борелевских множеств, 2) неотрицательная: $\Phi(e) \geq 0$, 3) абсолютно-аддитивная: если $e = e_1 + e_2 + \cdots$; $e_i \cap e_k = \emptyset$ $(i \neq k)$, то $\Phi(e) = \Phi(e_1) + \Phi(e_2) + \cdots$. Пусть $\Phi'(e')$ другое распределение масс, причем $\Phi(R) = \Phi'(R)$. Перемещением масс будем называть такую функцию $\Psi(e, e')$, определенную для пар (B)-совокупностей $e, e' \in R$: 1) неотрицательную и абсолютно-аддитивную по каждому из аргументов, 2) такую, что $\Psi(e, R) = \Phi(e)$; $\Psi(R, e') = \Phi'(e')$.

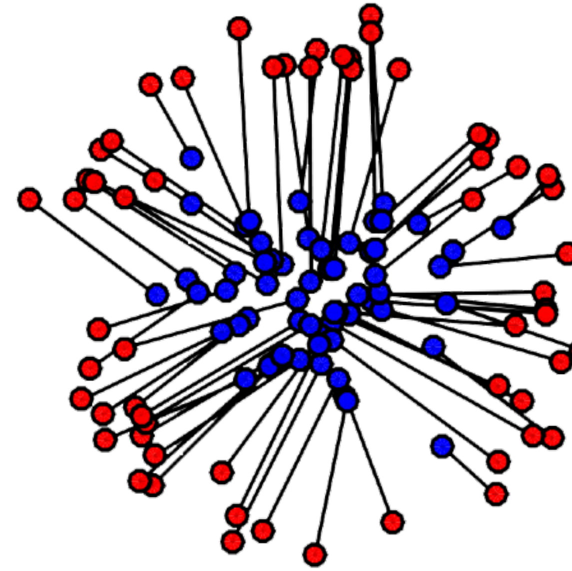
Пусть $r(x, y)$ известная непрерывная неотрицательная функция — работа по перемещению единицы массы из x в y.

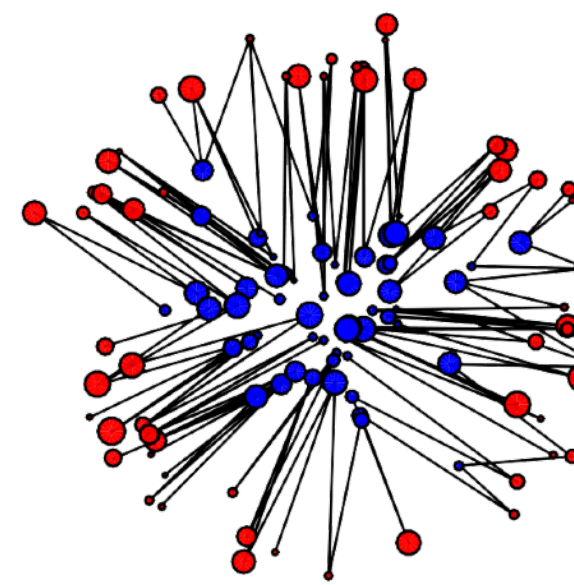
Работой по перемещению данных распределений масс будем называть величину

$$W(\Psi, \Phi, \Phi') = \int_R \int_R r(x, x')\Psi(de, de') = \lim_{\lambda \to 0} \sum_{i,k} r(x_i, x'_k)\Psi(e_i, e'_k),$$

где $\{e_i\}$ дизъюнктны и $\sum_1^n e_i = R$, $\{e'_k\}$ дизъюнктны и $\sum_1^m e'_k = R$, $x_i \in e_i$, $x'_k \in e'_k$, и λ наибольшее из чисел $\operatorname{diam} e_i$ $(i = 1, 2, \ldots, n)$, $\operatorname{diam} e'_k$ $(k = 1, 2, \ldots, m)$.

[1942]

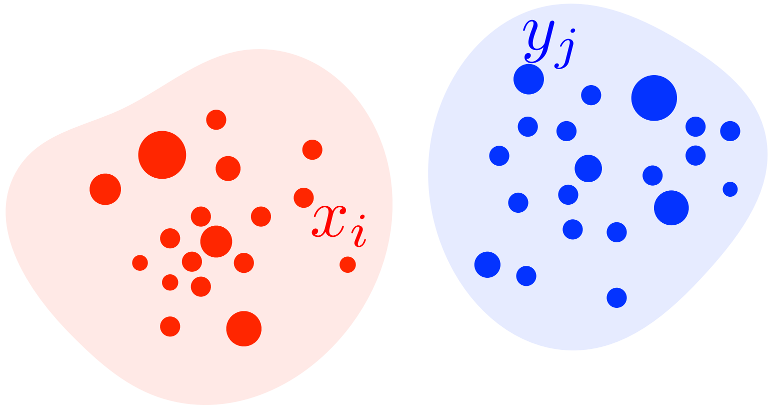




Kantorovitch's Formulation

Input distributions

$\alpha = \sum_{i=1}^n \mathbf{a}_i \delta_{x_i} \quad \beta = \sum_{j=1}^m \mathbf{b}_j \delta_{y_j}$



Points $(x_i)_i$, $(y_j)_j$

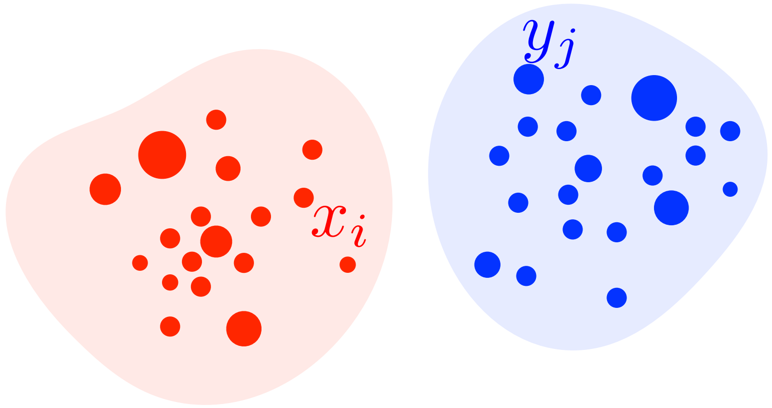
Weights $\mathbf{a}_i \geqslant 0$, $\mathbf{b}_j \geqslant 0$.

$\sum_{i=1}^n \mathbf{a}_i = \sum_{j=1}^m \mathbf{b}_j = 1$

Kantorovitch's Formulation

Input distributions

$\alpha = \sum_{i=1}^{n} \mathbf{a}_i \delta_{x_i} \quad \beta = \sum_{j=1}^{m} \mathbf{b}_j \delta_{y_j}$



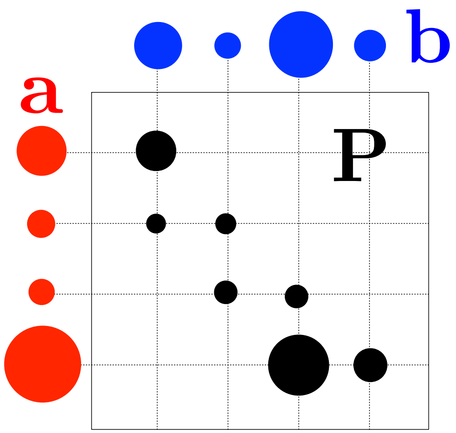
Points $(x_i)_i$, $(y_j)_j$

Weights $\mathbf{a}_i \geqslant 0$, $\mathbf{b}_j \geqslant 0$.

$\sum_{i=1}^{n} \mathbf{a}_i = \sum_{j=1}^{m} \mathbf{b}_j = 1$

Couplings:

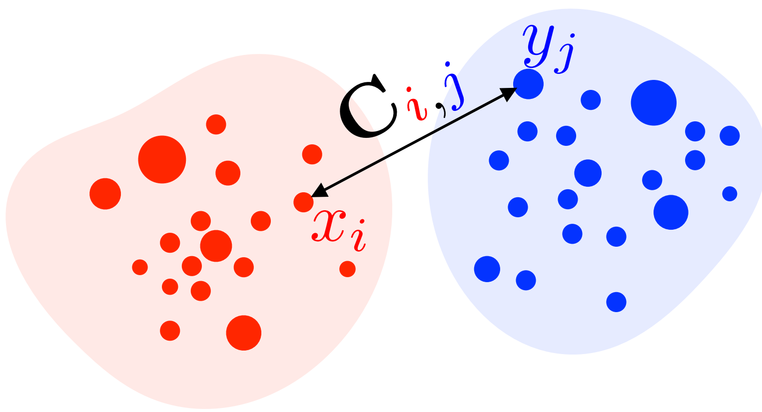
$$\mathbf{U}(\mathbf{a},\mathbf{b}) \overset{\text{def.}}{=} \left\{ \mathbf{P} \in \mathbb{R}_+^{n\times m} \; ; \; \mathbf{P}\mathbb{1}_n = \mathbf{a}, \mathbf{P}^\top \mathbb{1}_m = \mathbf{b} \right\}$$



Kantorovitch's Formulation

Input distributions

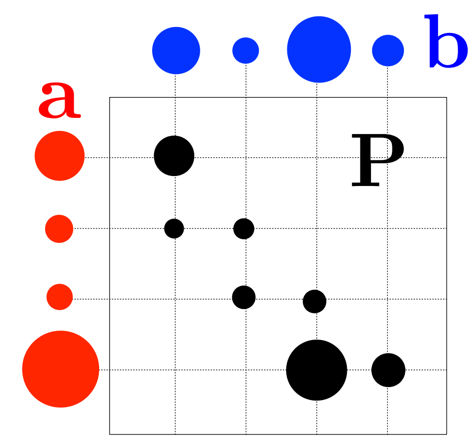
$\alpha = \sum_{i=1}^{n} \mathbf{a}_i \delta_{x_i} \quad \beta = \sum_{j=1}^{m} \mathbf{b}_j \delta_{y_j}$



Points $(x_i)_i$, $(y_j)_j$

Weights $\mathbf{a}_i \geqslant 0$, $\mathbf{b}_j \geqslant 0$.

$\sum_{i=1}^{n} \mathbf{a}_i = \sum_{j=1}^{m} \mathbf{b}_j = 1$



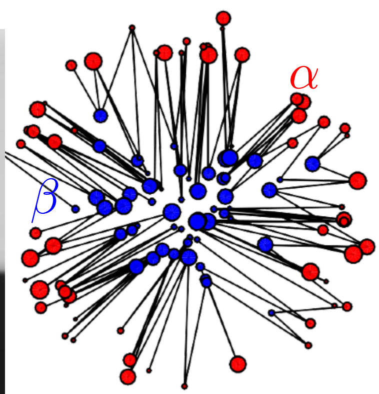
Couplings:

$$\mathbf{U}(\mathbf{a},\mathbf{b}) \stackrel{\text{def.}}{=} \left\{ \mathbf{P} \in \mathbb{R}_+^{n\times m} \; ; \; \mathbf{P}\mathbb{1}_n = \mathbf{a}, \mathbf{P}^\top \mathbb{1}_m = \mathbf{b} \right\}$$

Cost: $\quad \mathbf{C}_{i,j} = c(x_i, y_j)$

[Kantorovich 1942]

$$L_{\mathbf{C}}(\mathbf{a},\mathbf{b}) \stackrel{\text{def.}}{=} \min \left\{ \sum_{i,j} \mathbf{P}_{i,j} \mathbf{C}_{i,j} \; ; \; \mathbf{P} \in \mathbf{U}(\mathbf{a},\mathbf{b}) \right\}$$

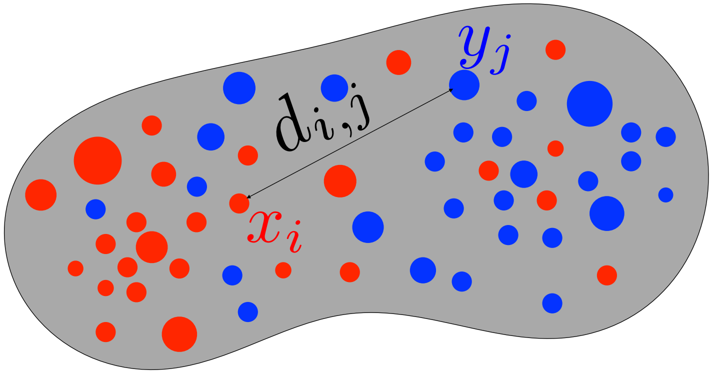


Wasserstein Distance

Metric spaces $\mathcal{X} = \mathcal{Y}$

Distance $d(x, y)$.

Cost $c(x, y) = d(x, y)^p$, $p \geqslant 1$.



Wasserstein distance:

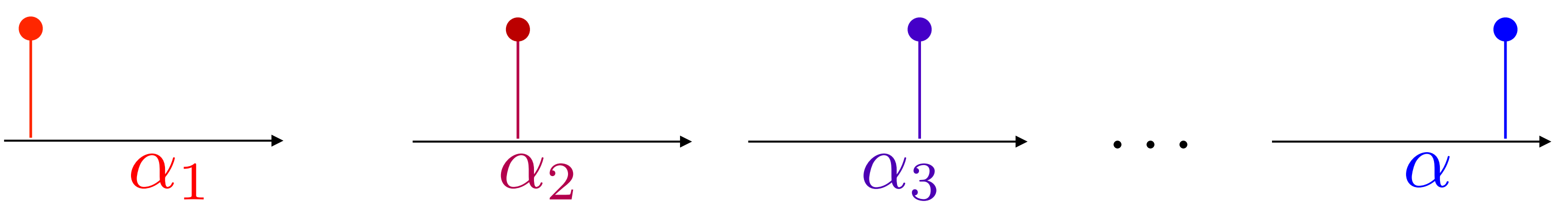
$\mathrm{W}_p(\mathbf{a}, \mathbf{b}) \stackrel{\text{def.}}{=} \mathrm{L}_{\mathbf{D}^p}(\mathbf{a}, \mathbf{b})^{1/p}$

$\mathcal{W}_p(\alpha, \beta) \stackrel{\text{def.}}{=} \mathcal{L}_{d^p}(\alpha, \beta)^{1/p}$

Theorem: W_p and $\mathcal{W}_p$ are distances.

$$\mathcal{W}_p(\alpha_n, \alpha) \rightarrow 0 \iff \alpha_n \overset{\text{weak}*}{\rightarrow} \alpha$$

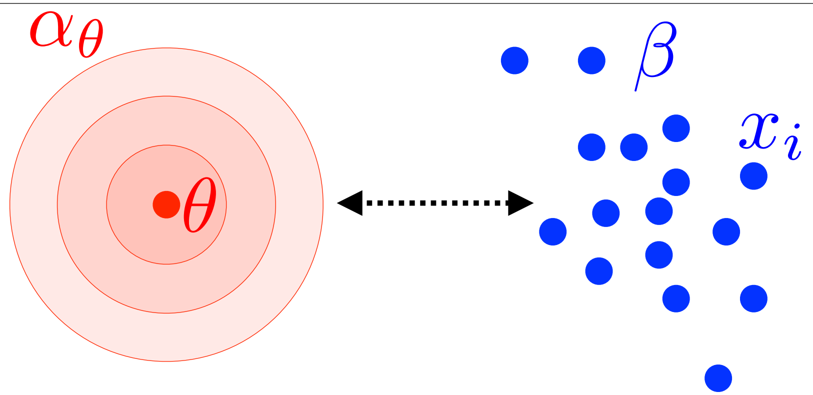
Examples: $\mathcal{W}_p(\delta_x, \delta_y) = d(x, y)$.



Density Fitting and Generative Models

Observations: $\beta \stackrel{\text{def.}}{=} \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$

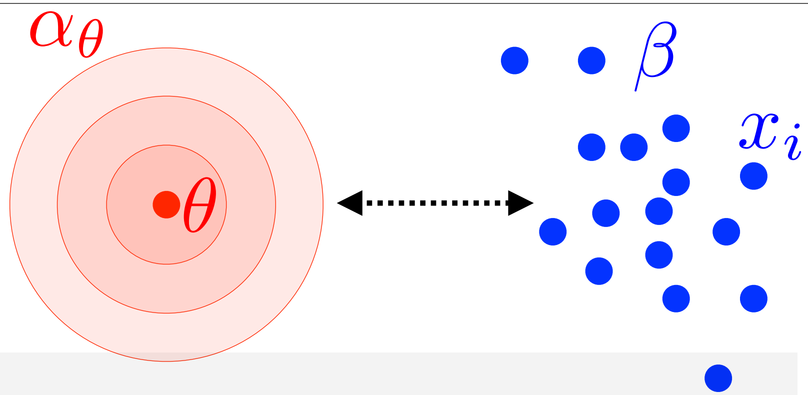
Parametric model: $\theta \mapsto \alpha_\theta$



Density Fitting and Generative Models

Observations: $\beta \stackrel{\text{def.}}{=} \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$

Parametric model: $\theta \mapsto \alpha_\theta$



Density fitting: $\mathrm{d}\alpha_\theta(x) = \rho_\theta(x)\mathrm{d}x$

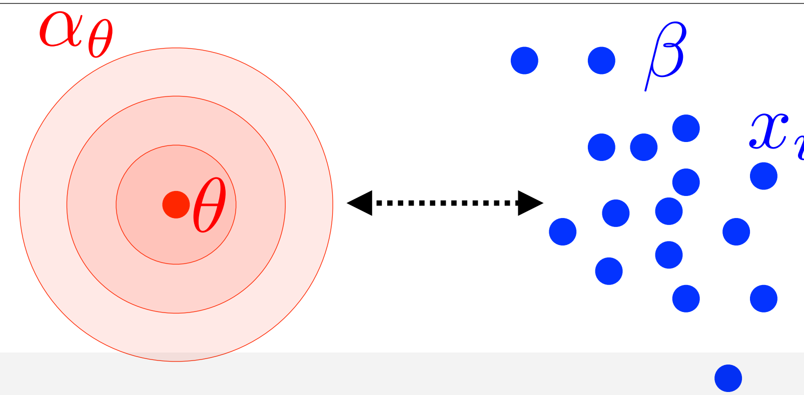
$$\min_\theta \widehat{\mathrm{KL}}(\beta|\alpha_\theta) \stackrel{\text{def.}}{=} -\sum_i \log(\rho_\theta(x_i))$$

Maximum likelihood (MLE)

Density Fitting and Generative Models

Observations: $\beta \overset{\text{def.}}{=} \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$

Parametric model: $\theta \mapsto \alpha_\theta$



Density fitting: $\mathrm{d}\alpha_\theta(x) = \rho_\theta(x)\mathrm{d}x$

$$\min_\theta \widehat{\mathrm{KL}}(\beta|\alpha_\theta) \overset{\text{def.}}{=} -\sum_i \log(\rho_\theta(x_i))$$

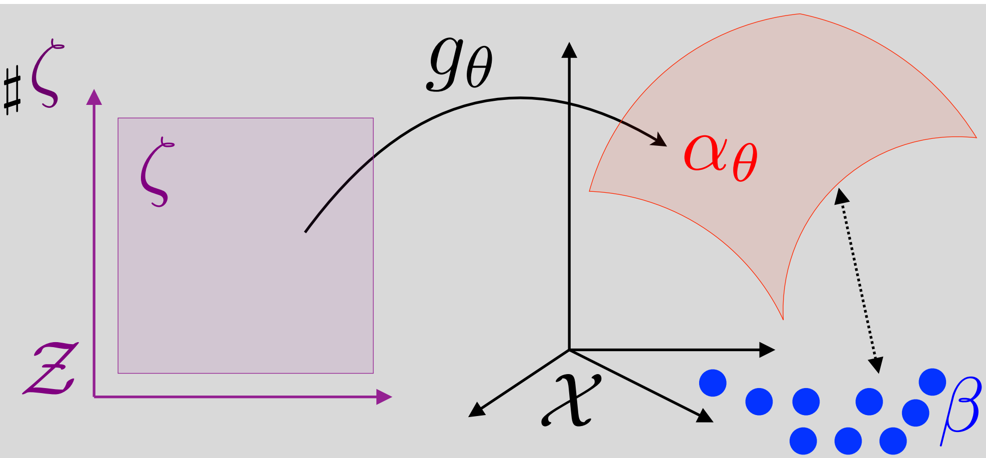
Maximum likelihood (MLE)

Generative model fit: $\alpha_\theta = g_{\theta,\sharp}\zeta$

$$\widehat{\mathrm{KL}}(\beta|\alpha_\theta) = +\infty$$

$\rightarrow$ MLE undefined.

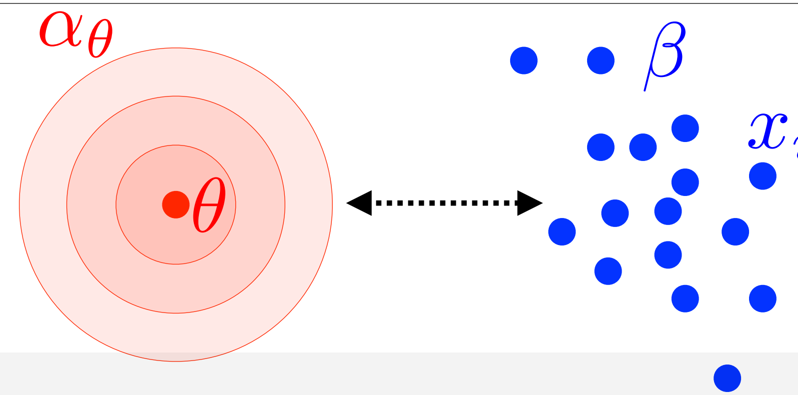
$\rightarrow$ Need a weaker metric.



Density Fitting and Generative Models

Observations: $\beta \overset{\text{def.}}{=} \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}$

Parametric model: $\theta \mapsto \alpha_\theta$



Density fitting: $\mathrm{d}\alpha_\theta(x) = \rho_\theta(x)\mathrm{d}x$

$$\min_\theta \widehat{\mathrm{KL}}(\beta|\alpha_\theta) \overset{\text{def.}}{=} -\sum_i \log(\rho_\theta(x_i))$$

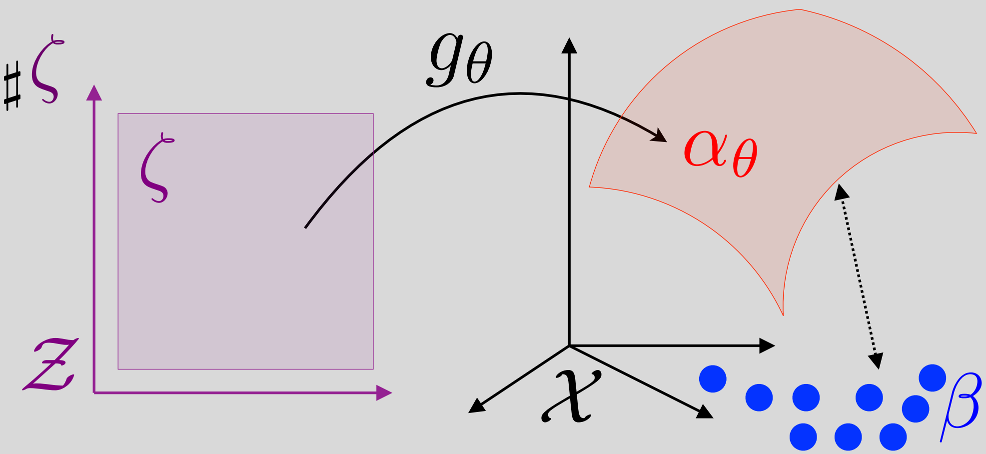
Maximum likelihood (MLE)

Generative model fit: $\alpha_\theta = g_{\theta,\sharp}\zeta$

$$\widehat{\mathrm{KL}}(\beta|\alpha_\theta) = +\infty$$

$\rightarrow$ MLE undefined.

$\rightarrow$ Need a weaker metric.



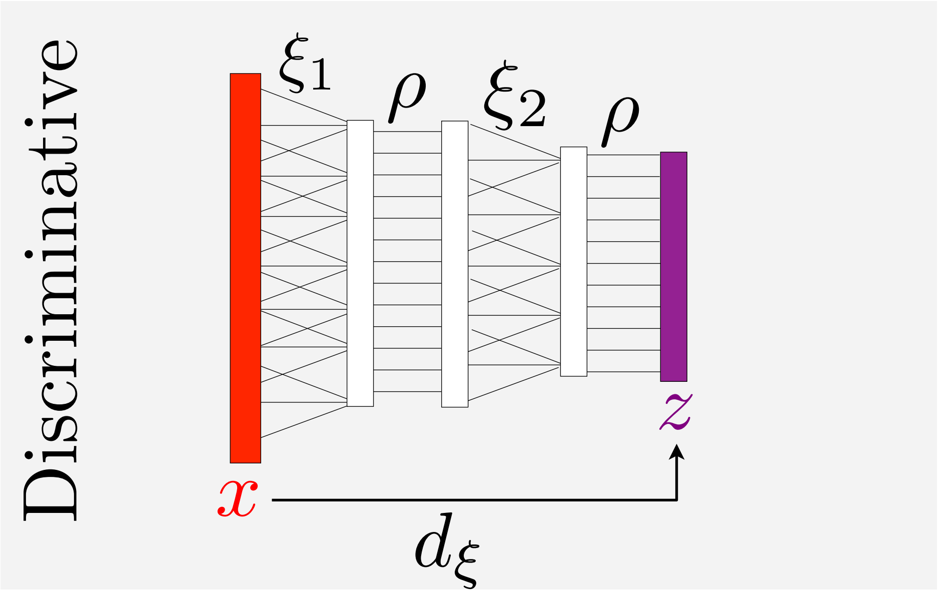
Proposal: Wasserstein fitting $\min_\theta \mathcal{W}_p(\alpha_\theta, \beta)$

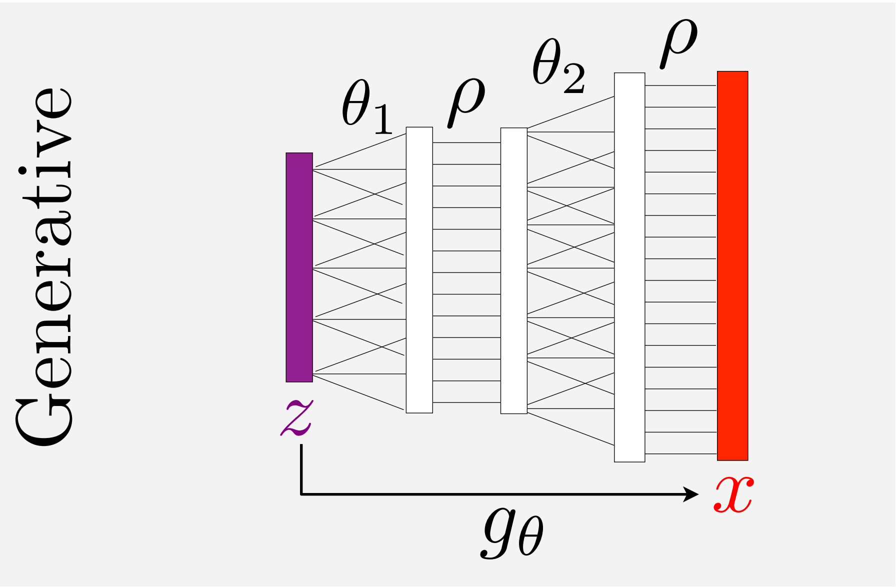
Deep Discriminative vs Generative Models

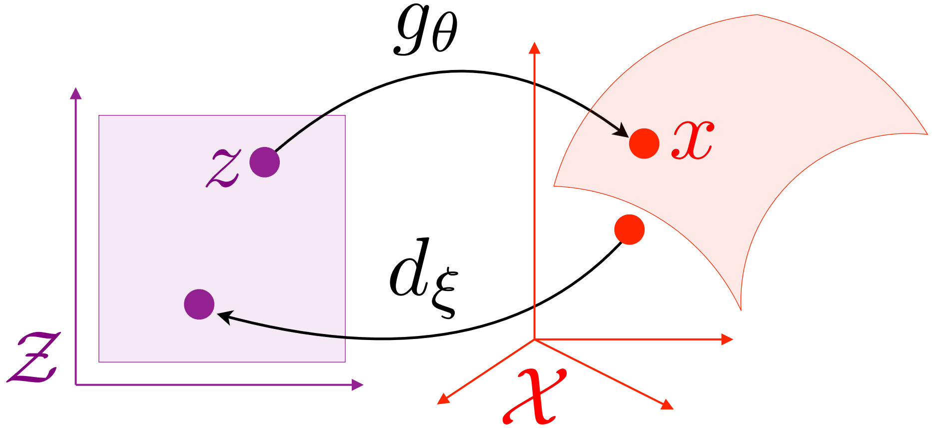
Deep networks:

$$d_\xi(x) = \rho(\xi_K(\ldots\rho(\xi_2(\rho(\xi_1(x))\ldots)$$

$$g_\theta(z) = \rho(\theta_K(\ldots\rho(\theta_2(\rho(\theta_1(z))\ldots)$$





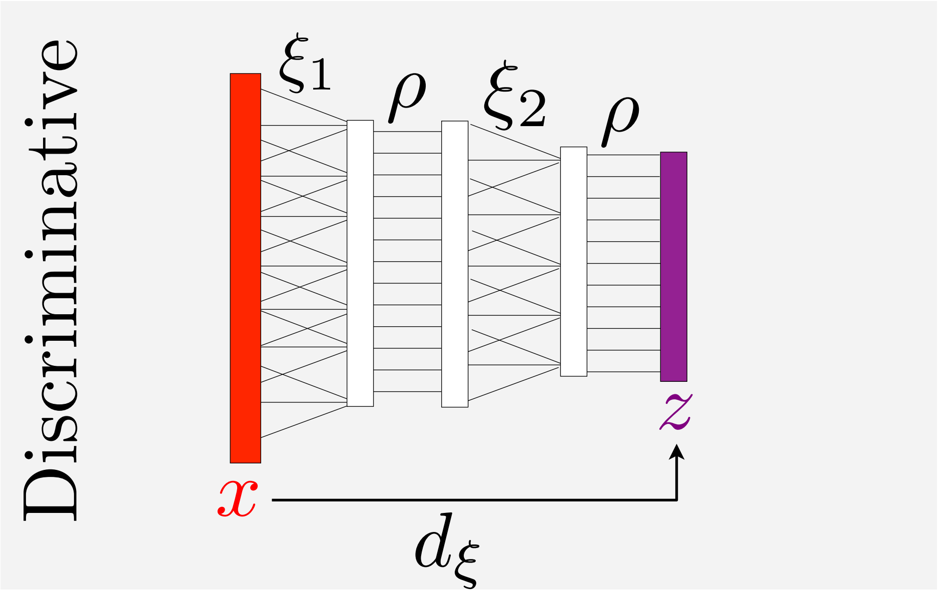


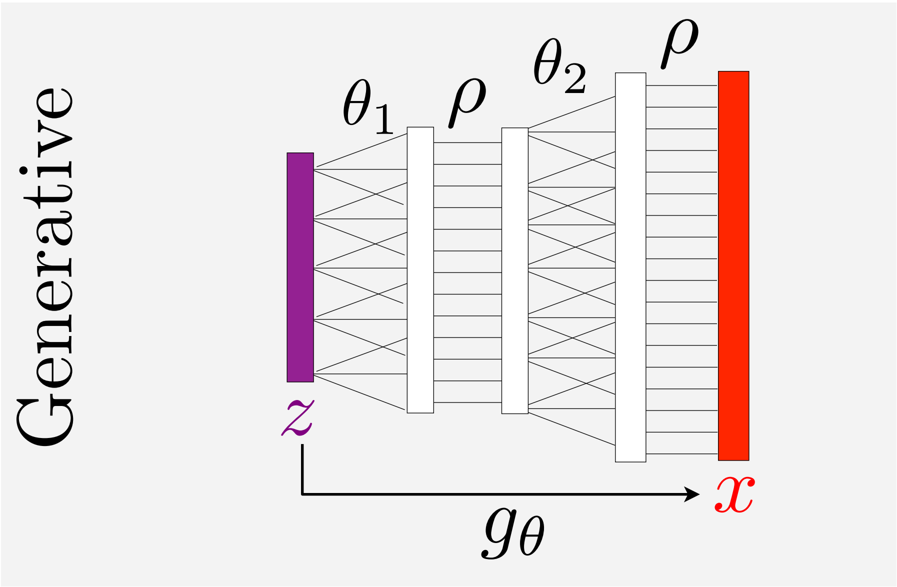
Deep Discriminative vs Generative Models

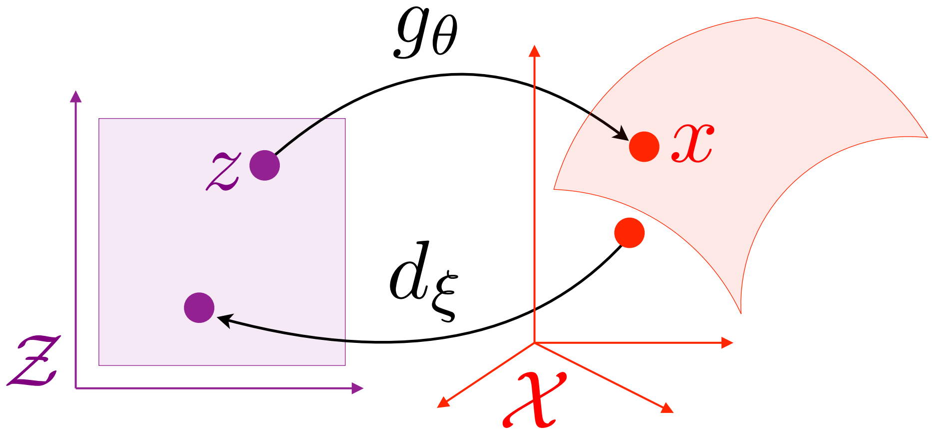
Deep networks:

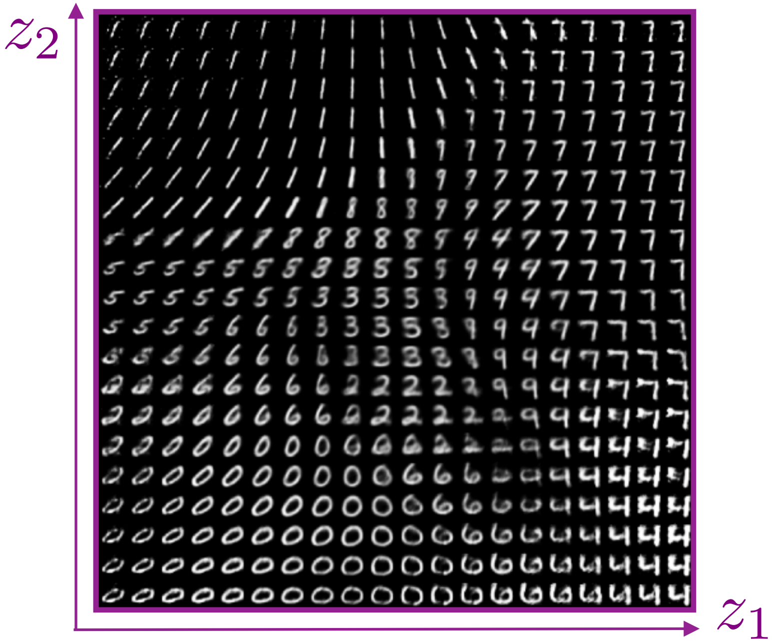
$$d_\xi(x) = \rho(\xi_K(\ldots\rho(\xi_2(\rho(\xi_1(x))\ldots)$$

$$g_\theta(z) = \rho(\theta_K(\ldots\rho(\theta_2(\rho(\theta_1(z))\ldots)$$

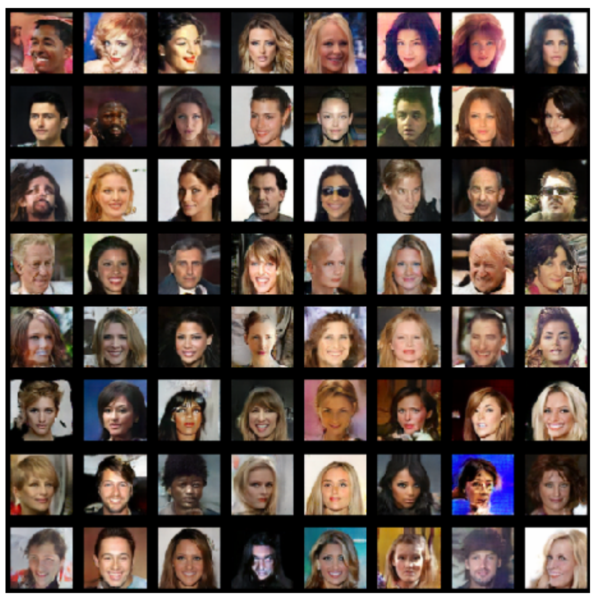


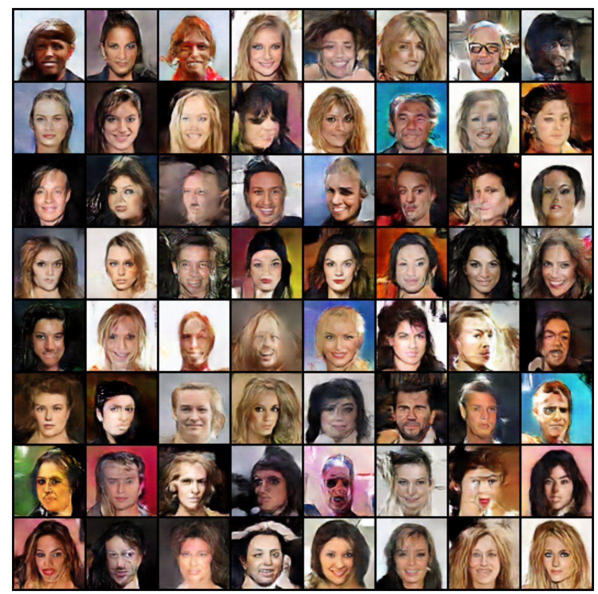






Examples of Image Generation





Inputs

→ Need to learn the metric $c(x,y) = \|d_\xi(x) - d_\xi(y)\|^p$ ($\sim$ GANs)

→ Performance evaluation of generative models is an open problem.

Progressive Growing of GANs for Improved Quality, Stability, and Variation
Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen, ICLR 2018

Progressive Growing of GANs for Improved Quality, Stability, and Variation
Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen, ICLR 2018

Conclusion: Toward High-dimensional OT

Monge Kantorovich Dantzig Brenier Otto McCann Villani Figalli