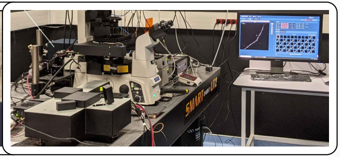
Supervised Learning & Inverse Problems in Imaging
(or How I Became a Bayesian)

Pierre Weiss, CNRS, Université de Toulouse

Introduction

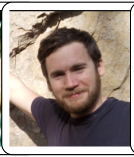
Can we recover the missing information?

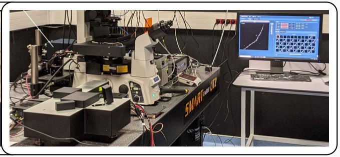
The RIM team philosophy



- "If we cannot measure, we should change the acquisition protocol"

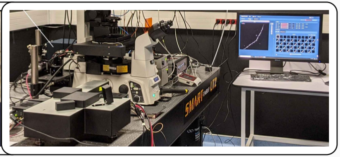
The RIM team philosophy





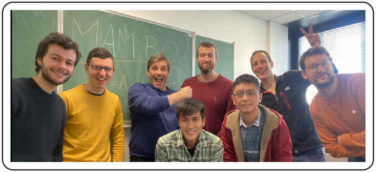
- “If we cannot measure, we should change the acquisition protocol”
- “You can trust every pixel we provide”

The RIM team philosophy



- “If we cannot measure, we should change the acquisition protocol”
- “You can trust every pixel we provide”

The deviant approach (today)



- “If you don’t have the information just invent it!”

Asking my students

Asking my students

Asking my students

Asking my students

Asking my students

Asking my students

Asking my students

Observation

Observation

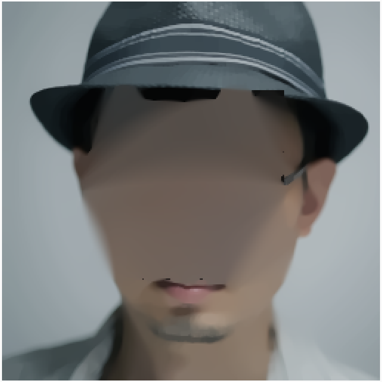
- Average scientist scholar $\neq$ fine arts school student
- Not much can be said... Likely 2 eyes, glasses and a nose!

Observation

- Average scientist scholar $\neq$ fine arts school student
- Not much can be said... Likely 2 eyes, glasses and a nose!

Main objectives today

- Can computers can do this better?
- What are the main mathematical tools and their properties?
- Some practical implications

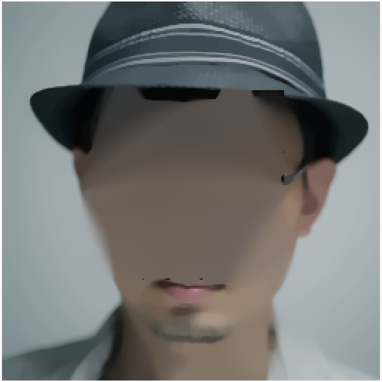


Nonlinear **total variation** based noise removal algorithms
LI Rudin, S Osher, E Fatemi - Physica D: nonlinear phenomena, 1992 - Elsevier
A constrained optimization type of numerical algorithm for removing noise from images is presented. The **total variation** of the image is minimized subject to constraints involving the …
☆ Enregistrer 99 Citer Cité 19060 fois Autres articles Les 24 versions Web of Science: 10580

1990-2015: "handcrafted" approaches

Total variation minimization

$$\hat{x} = \underset{x|_{\text{mask}}=y|_{\text{mask}}}{\operatorname{argmin}} \|\nabla x\|_1$$



Nonlinear **total variation** based noise removal algorithms
LI Rudin, S Osher, E Fatemi - Physica D: nonlinear phenomena, 1992 - Elsevier
A constrained optimization type of numerical algorithm for removing noise from images is presented. The **total variation** of the image is minimized subject to constraints involving the …
☆ Enregistrer 99 Citer Cité 19060 fois Autres articles Les 24 versions Web of Science: 10580

1990-2015: "handcrafted" approaches

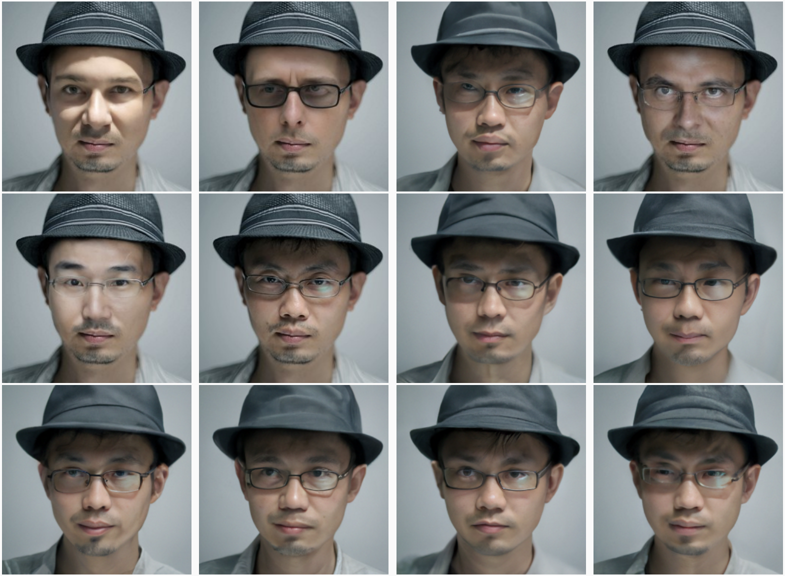
Total variation minimization

$$\hat{x} = \underset{x|_{\text{mask}}=y|_{\text{mask}}}{\operatorname{argmin}} \|\nabla x\|_1$$

Limited practical interest!

Today: AI and the MAP. $\hat{x} = \mathrm{argmax}_x\, p(x|y)$

Today: AI and the MMSE. $\hat{x} = \mathbb{E}(x|y)$



Today: AI and sampling the posterior $p_{x|y}$

Preliminaries

Inverse problems

$$y = A(x) + b$$

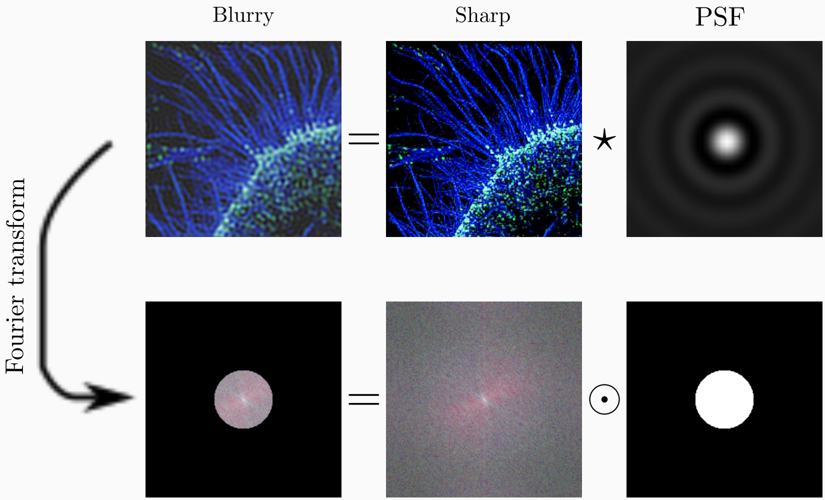
- $A : \mathbb{R}^N \to \mathbb{R}^M$ observation operator
- x: image to recover
- b: noise
- y: observed measurements

Inverse problem $\equiv$ recover x from y

Inpainting

A = Mask Multiplication (inpainting in the space domain)

Microscopy super-resolution



Loss of high frequencies!

A = Convolution = Fourier + Mask (outpainting in the Fourier domain)

Magnetic Resonance Imaging

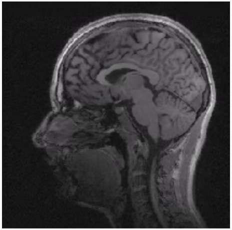
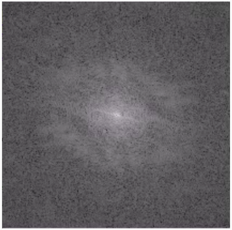
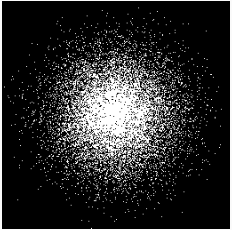


Image x



Fourier transform



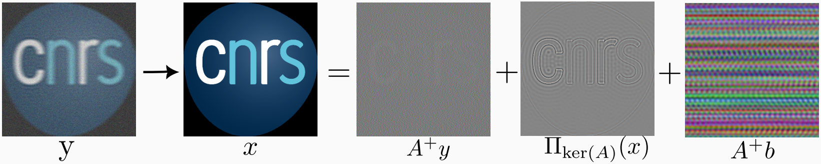
Sampling scheme

A = Fourier + Mask (inpainting in the Fourier domain)

The main difficulties in linear inverse problems

$$A^{-1}(y) = \{A^{+}y + \ker(A) + A^{+}b\}$$

- $A^{+}y$: information "available" on x
- $\ker(A)$: information lost in the process
- $A^{+}b$: structured noise



Bayesian formalism

Some information is lost in the acquisition!

We inject it through a probabilistic model.

- x is the realization of a random variable X with density p_X.
- b is the realization of a random variable B with density p_B.

We'll write $p(x)$ for $p_X(x)$ to simplify the notation.

What is a prior p_X?

A function that evaluates the probability of an image.

If X is the image of a sheep.

$p_X($  $)$

What is a prior p_X?

A function that evaluates the probability of an image.

If X is the image of a sheep.

p_X () $\geq p_X$ ()

What is a prior p_X?

A function that evaluates the probability of an image.

If X is the image of a sheep.

$p_X($  $) \geq p_X($  $) \geq p_X($  $)$

What is a prior p_X?

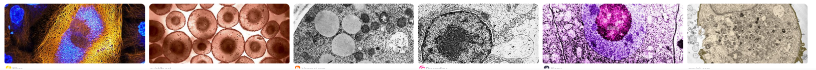
A function that evaluates the probability of an image.

If X is the image of a sheep.

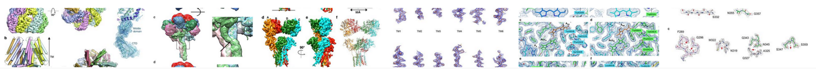
$p_X\Bigg($  $\Bigg) \geq p_X\Bigg($  $\Bigg) \geq p_X\Bigg($  $\Bigg) \geq p_X\Bigg($  $\Bigg)$ $= 0$!

What is a prior p_X?

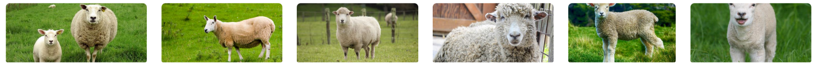
- It is domain specific
- More or less spread depending on the amount of information



Sampling $p_{X|\text{cell}}$



Sampling $p_{X|\text{molecule}}$



Sampling $p_{X|\text{sheep}}$

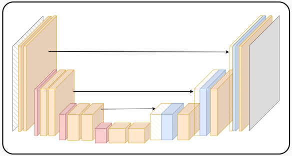
MAP, MMSE and Posterior Sampling

Learning to denoise

- $A = \mathrm{Id}$ and $B \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$
- $Y = X + B$: noisy image.

Learning to denoise

- $A = \mathrm{Id}$ and $B \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$
- $Y = X + B$: noisy image.

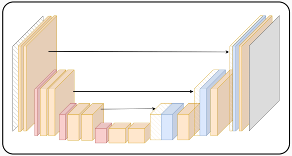


Prerequisities

- Neural network $N(y, w)$.
- A database of clean images $(x_1, \ldots, x_I)$
- Synthesize $y_i = x_i + b_i$.

Learning to denoise

- $A = \mathrm{Id}$ and $B \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$
- $Y = X + B$: noisy image.



Prerequisities

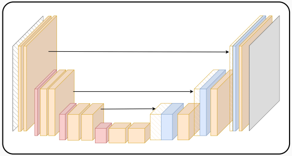
- Neural network $N(y, w)$.
- A database of clean images $(x_1, \ldots, x_I)$
- Synthesize $y_i = x_i + b_i$.

Training $\equiv$ Stochastic gradient

- $\inf_w \frac{1}{I} \sum_{i=1}^{I} \|N(y_i, w) - x_i\|_2^2$

Learning to denoise

- $A = \mathrm{Id}$ and $B \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$
- $Y = X + B$: noisy image.



Prerequisities

- Neural network $N(y, w)$.
- A database of clean images $(x_1, \ldots, x_I)$
- Synthesize $y_i = x_i + b_i$.

Training $\equiv$ Stochastic gradient

- $\inf_w \frac{1}{I} \sum_{i=1}^{I} \|N(y_i, w) - x_i\|_2^2$

Output

- $N(y, w^\star)$: a trained network
- Can be used with arbitrary images

Minimum Mean Square Estimation (MMSE)

$$\hat{x}_{MMSE}(y) \stackrel{\text{def.}}{=} \operatorname*{argmin}_{x \in \mathbb{R}^N} \mathbb{E}(\|x - X\|_2^2 | Y = y)$$

Minimum Mean Square Estimation (MMSE)

$$\hat{x}_{MMSE}(y) \overset{\textbf{def.}}{=} \operatorname*{argmin}_{x \in \mathbb{R}^N} \mathbb{E}(\|x - X\|_2^2 | Y = y) = \mathbb{E}(x|y)$$

Minimum Mean Square Estimation (MMSE)

$$\hat{x}_{MMSE}(y) \stackrel{\text{def.}}{=} \operatorname*{argmin}_{x \in \mathbb{R}^N} \mathbb{E}(\|x - X\|_2^2 | Y = y) = \mathbb{E}(x|y)$$

Claim (informal)

If I is large enough, $N(\cdot, w)$ is expressive and good training.

$$N(y, w^\star) \approx \hat{x}_{MMSE}(y)$$

Minimum Mean Square Estimation (MMSE)

$$\hat{x}_{MMSE}(y) \overset{\text{def.}}{=} \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \mathbb{E}(\|x - X\|_2^2 | Y = y) = \mathbb{E}(x|y)$$

Claim (informal)

If I is large enough, $N(\cdot, w)$ is expressive and good training.

$$N(y, w^\star) \approx \hat{x}_{MMSE}(y)$$

Proof.

$$\begin{aligned}
\hat{x}_{MMSE} : y \mapsto & \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \mathbb{E}(\|x - X\|_2^2 | Y = y) && \text{Average risk} \\
& \approx \underset{x : \mathbb{R}^N \to \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{I} \sum_{i=1}^{I} \|x(y_i) - x_i\|_2^2 && \text{Empirical risk} \\
& \approx \underset{w}{\operatorname{argmin}} \frac{1}{I} \sum_{i=1}^{I} \|N(y_i, w) - x_i\|_2^2 && \text{NN Approximation} \\
& \approx N(\cdot, w^\star)! && \text{Good optimization}
\end{aligned}$$

□

MMSE denoising $\approx$ prior

Assume that $Y = X + B$ with $B \sim \mathcal{N}(0, \delta^2 \mathrm{Id})$. Then

$$p_Y = p_X \star G_\delta \qquad \text{Basic property}$$

MMSE denoising $\approx$ prior

Assume that $Y = X + B$ with $B \sim \mathcal{N}(0, \delta^2 \mathrm{Id})$. Then

$$p_Y = p_X \star G_\delta$$ Basic property

$$\begin{aligned} \nabla \log p_Y(y) &= \frac{y - \hat{x}_{MMSE}(y)}{\delta^2} && \text{Tweedie Formula} \\ &\approx \frac{y - N(y, w^\star)}{\delta^2} && \text{NN power} \\ &\approx \nabla \log p_X(y) && \text{Small } \delta \end{aligned}$$

MMSE denoising $\approx$ prior

Assume that $Y = X + B$ with $B \sim \mathcal{N}(0, \delta^2 \mathrm{Id})$. Then

$$p_Y = p_X \star G_\delta \qquad \text{Basic property}$$

$$\begin{aligned} \nabla \log p_Y(y) &= \frac{y - \hat{x}_{MMSE}(y)}{\delta^2} && \text{Tweedie Formula} \\ &\approx \frac{y - N(y, w^\star)}{\delta^2} && \text{NN power} \\ &\approx \nabla \log p_X(y) && \text{Small } \delta \end{aligned}$$

Good denoiser $\approx$ gradient of the log prior!

Computing the MAP (Plug&Play prior)

Assume that $Y = A(X) + B$.

$$\hat{x}_{MAP}(y) \overset{\text{def.}}{=} \underset{x \in \mathbb{R}^N}{\operatorname{argmax}} \, p_{X|Y}(x|y)$$

Computing the MAP (Plug&Play prior)

Assume that $Y = A(X) + B$.

$$\hat{x}_{MAP}(y) \stackrel{\text{def.}}{=} \underset{x \in \mathbb{R}^N}{\operatorname{argmax}}\, p_{X|Y}(x|y)$$

Can be computed with a gradient descent:

$$\begin{aligned} x_{k+1} &= x_k + \tau \nabla p_{X|Y}(x_k|y) \\ &\stackrel{Bayes}{=} x_k - \tau \left[-\nabla \log p_{Y|X}(y|x_k) - \nabla \log p_X(x_k) \right] \\ &\stackrel{Tweedie}{\approx} x_k - \tau \left[-\nabla \log p_{Y|X}(y|x_k) - \frac{x_k - N(x_k, w^\star)}{\delta^2} \right]. \end{aligned}$$

Plug-and-play priors for model based reconstruction
SV **Venkatakrishnan**, CA Bouman… - 2013 IEEE global …, 2013 - ieeexplore.ieee.org
… This framework, which we term as **Plug**-and-**Play** priors, has the advantage that it … We demonstrate with some simple examples how **Plug**-and-**Play** priors can be used to mix and match …
☆ Enregistrer 99 Citer Cité 1014 fois Autres articles Les 14 versions

Bayesian imaging using plug & play priors: when langevin meets tweedie
R **Laumont**, VD Bortoli, A Almansa, J **Delon**… - SIAM Journal on Imaging …, 2022 - SIAM
Since the seminal work of Venkatakrishnan, Bouman, and Wohlberg [Proceedings of the Global Conference on Signal and Information Processing, IEEE, 2013, pp. 945--948] in 2013, …
☆ Enregistrer 99 Citer Cité 77 fois Autres articles Les 9 versions

Sampling the posterior

Assume that $Y = A(X) + B$. Construct the Euler-Maruyama sequence:

$$x_{k+1} = x_k - \tau \left[\nabla \log p_{X|Y}(x_k|y) + \sqrt{2} b_k \right]$$

where $b_k \sim \mathcal{N}(0, \mathrm{Id})$.

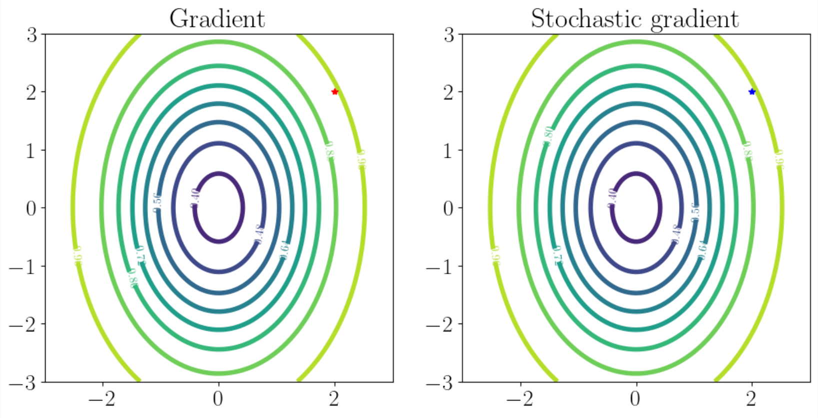
Then (under mild conditions – log-Sobolev inequalities)

$$\frac{1}{K} \sum_{k=1}^{K} \delta_{x_k} \rightharpoonup p_{X|Y}$$

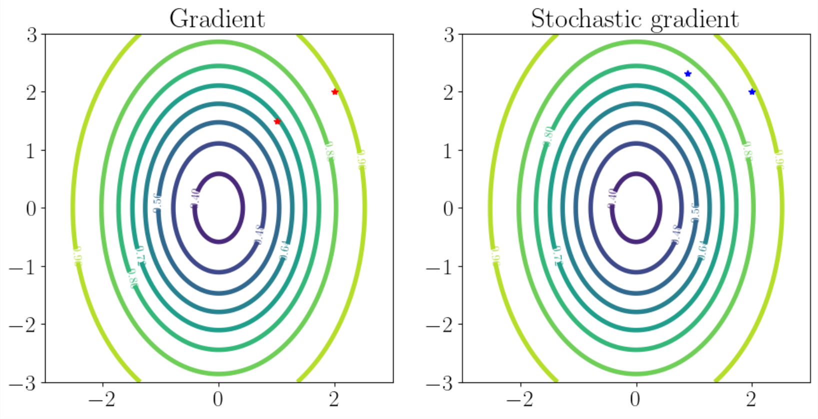
Denoising diffusion probabilistic models
J Ho, A Jain, P Abbeel - Advances in neural information …, 2020 - proceedings.neurips.cc
… This paper presents progress in **diffusion probabilistic models** [53]. A **diffusion probabilistic model** (which we will call a "**diffusion model**" for brevity) is a parameterized Markov chain …

Exponential convergence of Langevin distributions and their discrete approximations
GO Roberts, RL Tweedie - Bernoulli, 1996 - JSTOR
… **Convergence** of **Langevin** diffusions 345 We will see that **there** is **exponential convergence** of the m 7r E G9m. This behaviour is identical to that exhibited by the m algorithm, as shown …

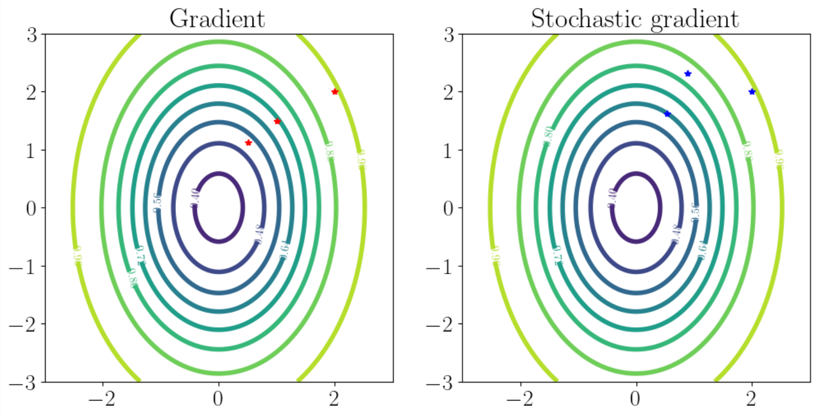
A starting problem



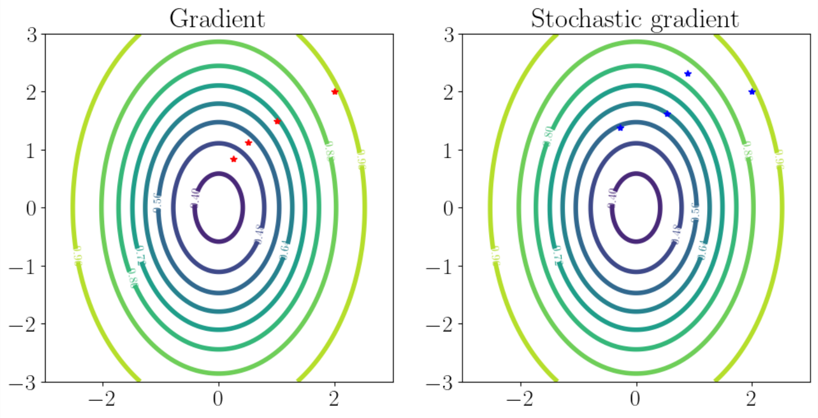
A starting problem



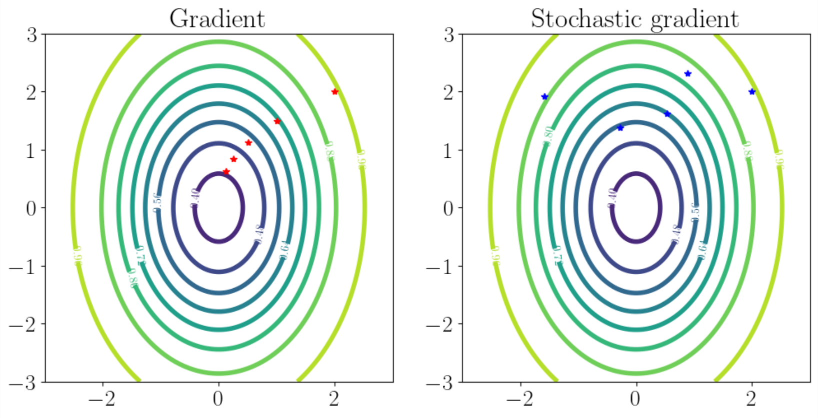
A starting problem



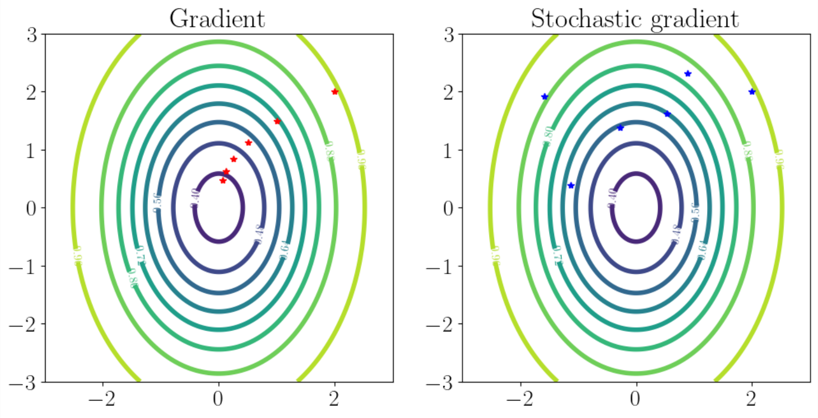
A starting problem



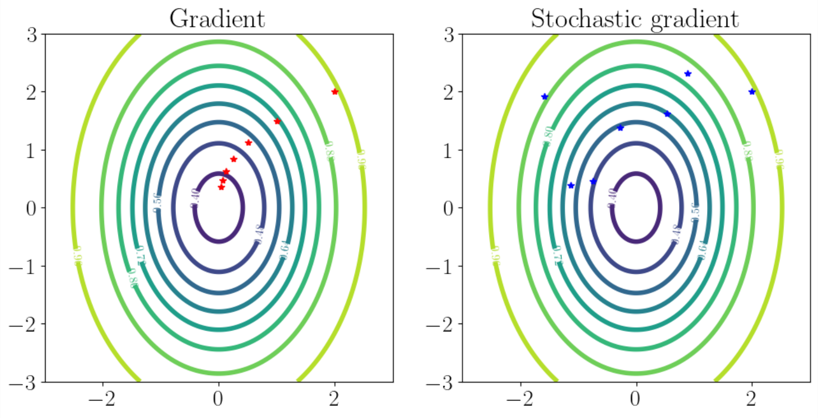
A starting problem



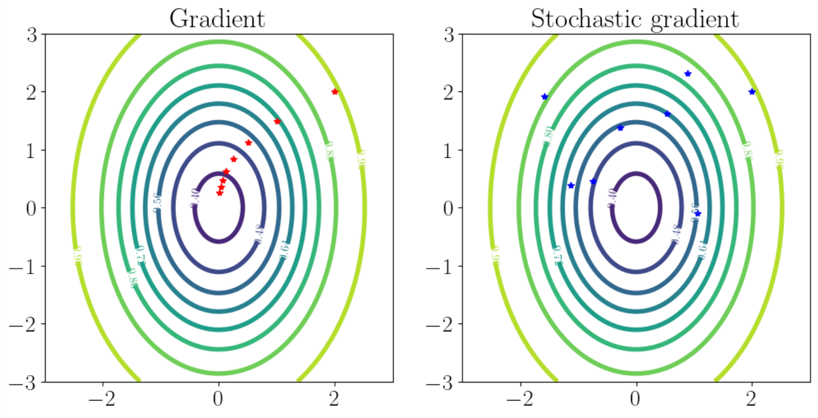
A starting problem



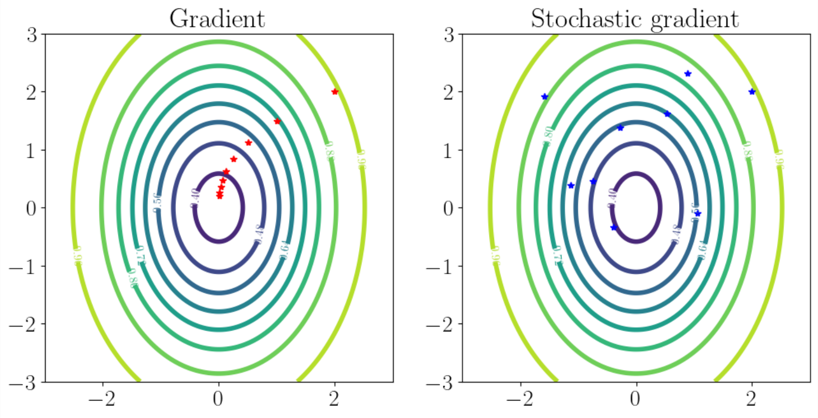
A starting problem



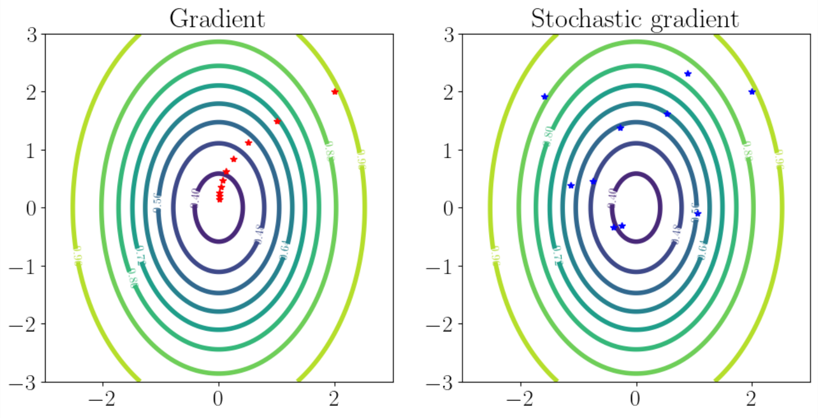
A starting problem



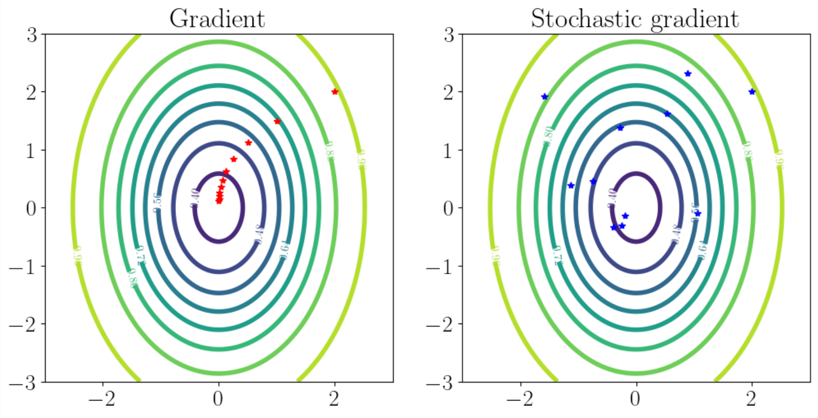
A starting problem



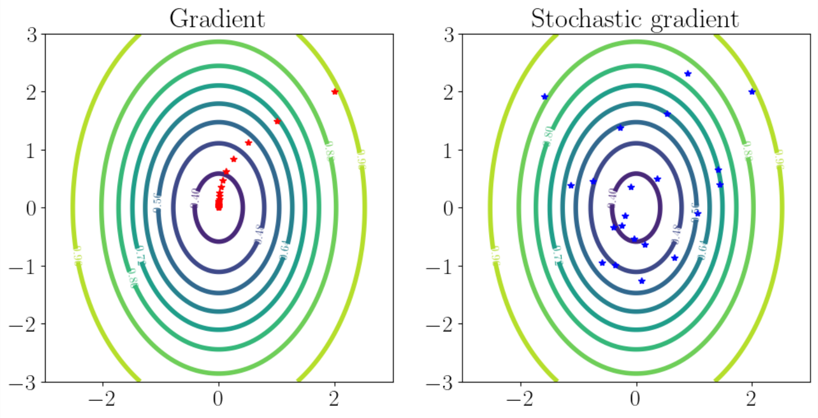
A starting problem



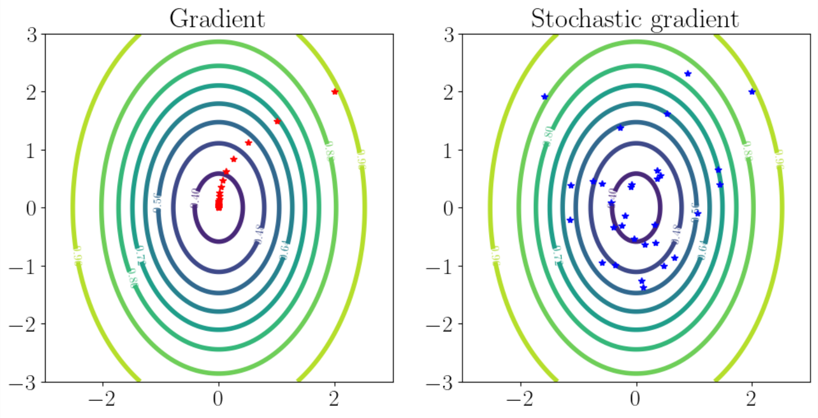
A starting problem



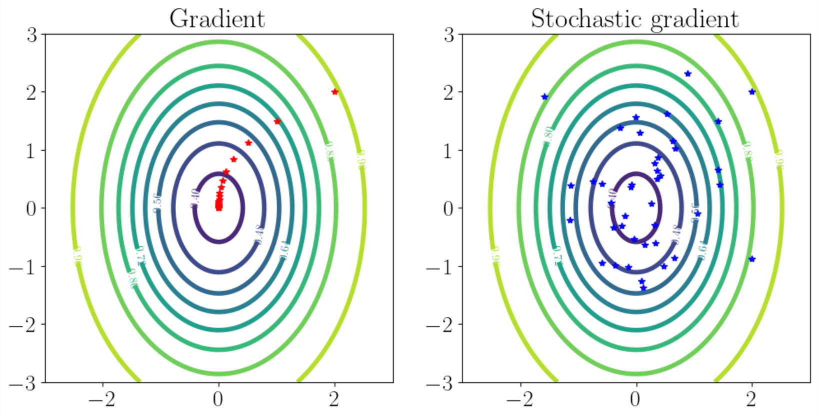
A starting problem



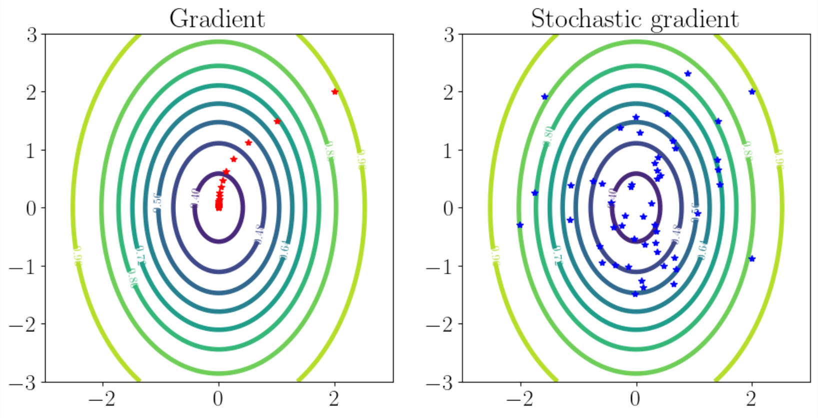
A starting problem



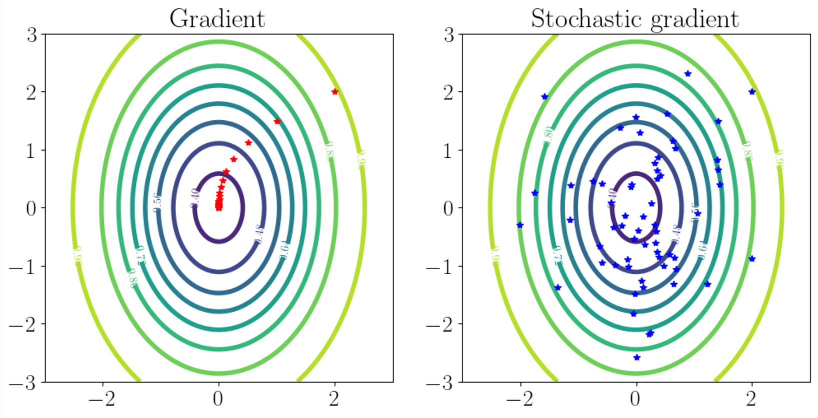
A starting problem

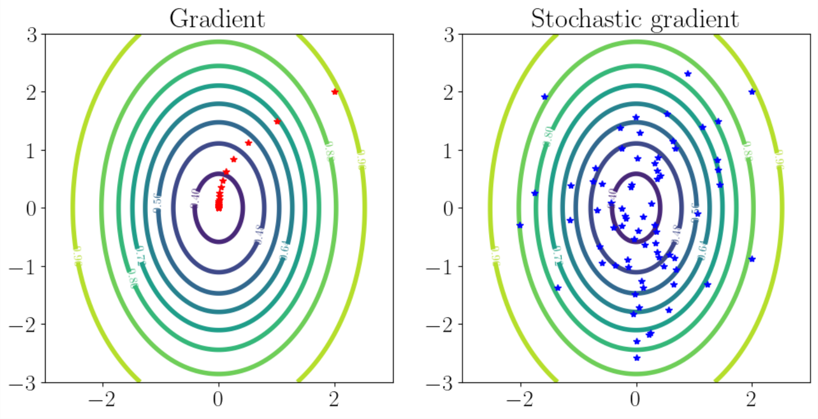


A starting problem

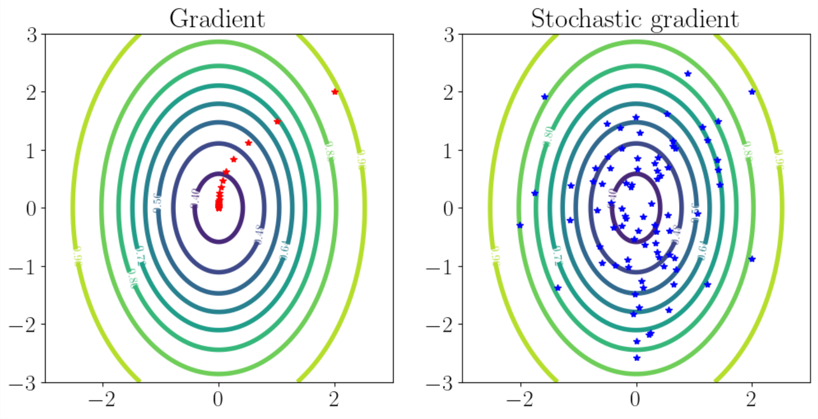


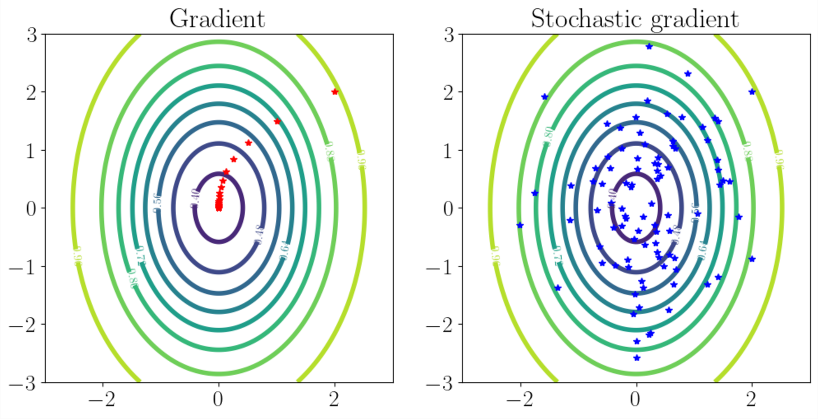
A starting problem



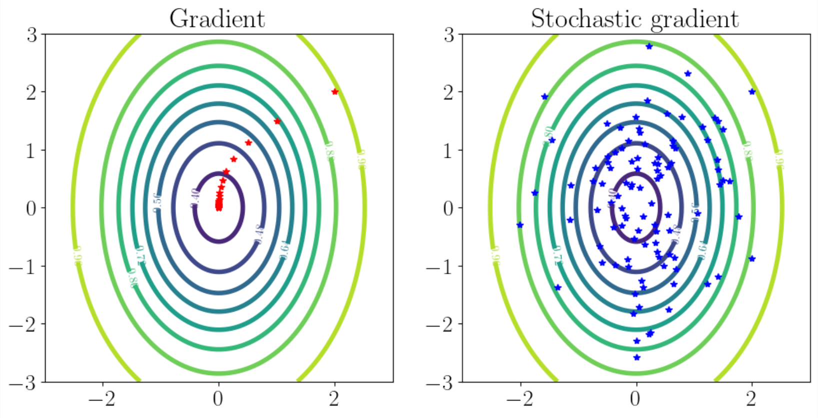

Gradient
Stochastic gradient
3
2
1
0
−1
−2
−3
−2
0
2

A starting problem

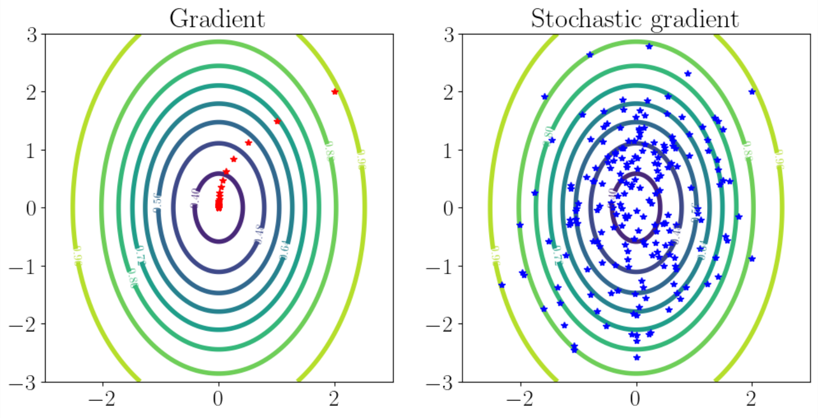



Gradient
Stochastic gradient
3
2
1
0
−1
−2
−3
−2
0
2

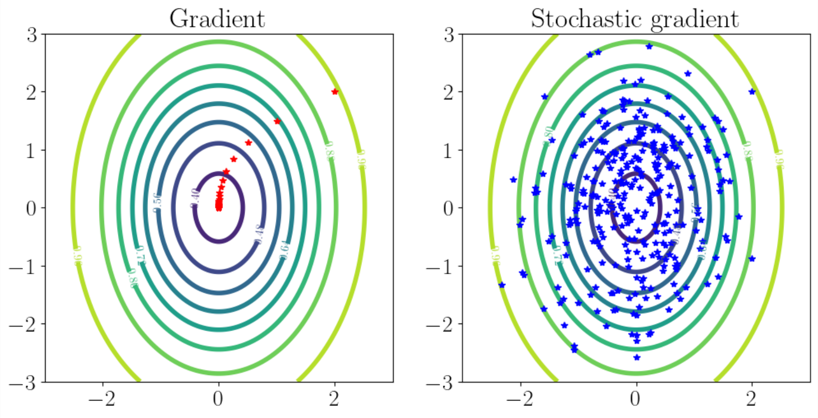
A starting problem



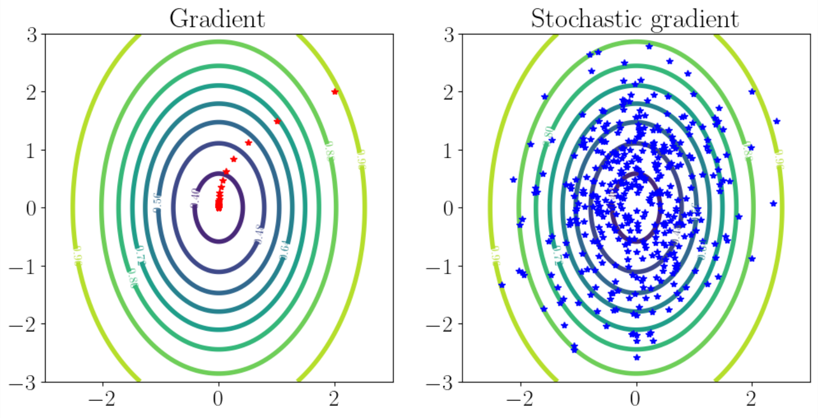
A starting problem

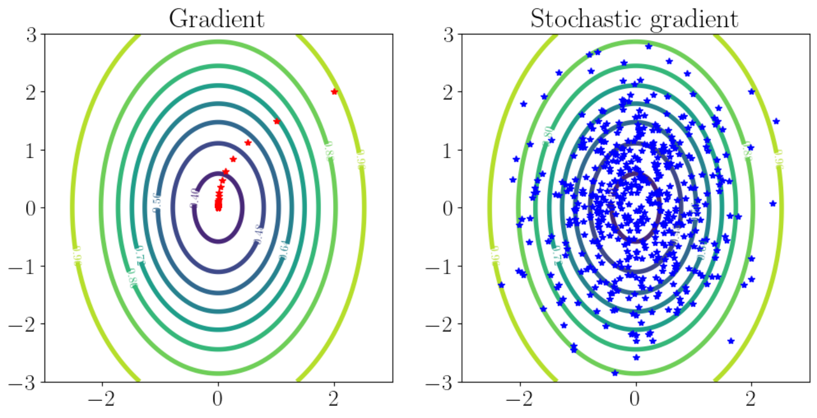


A starting problem

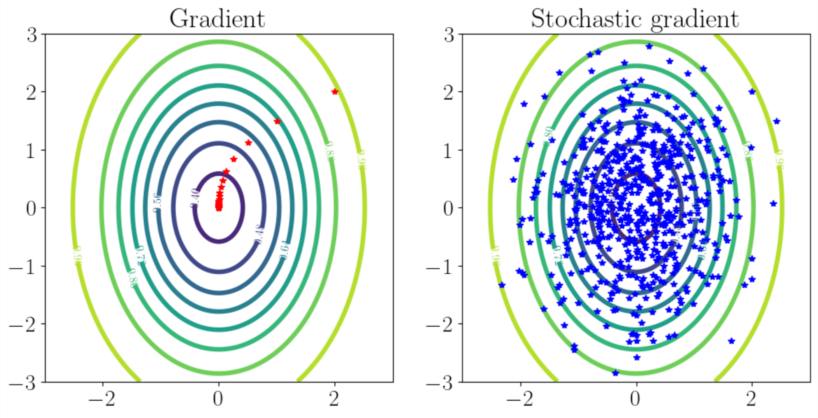


A starting problem

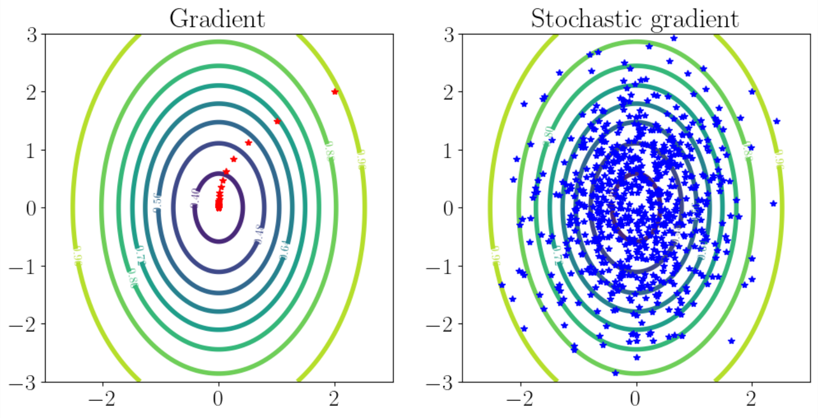


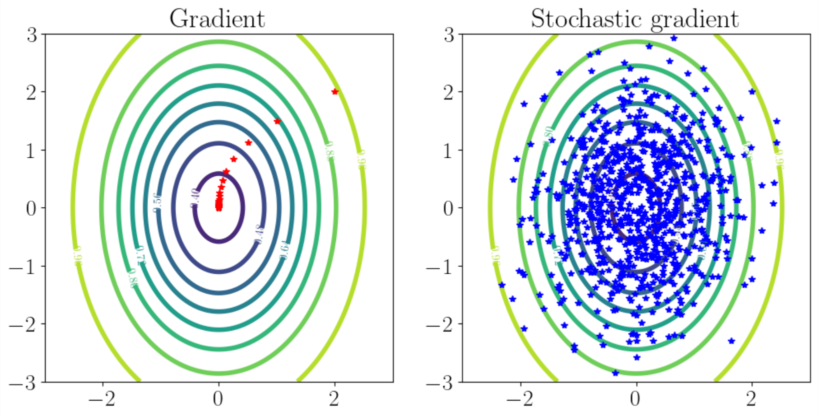

Gradient
Stochastic gradient
3
2
1
0
−1
−2
−3
−2
0
2

A starting problem

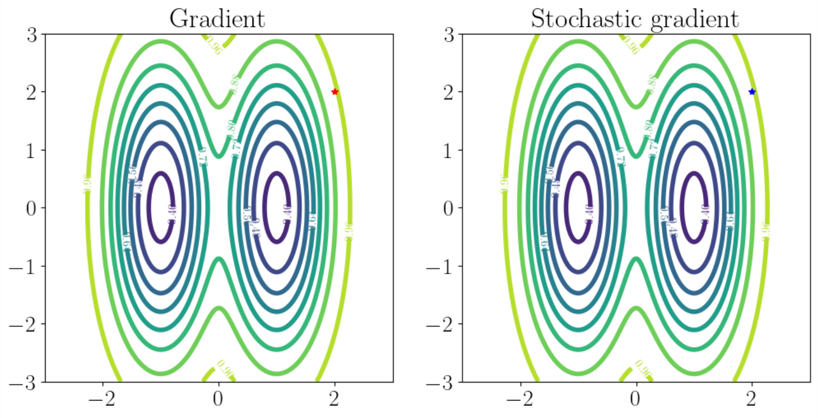


A starting problem

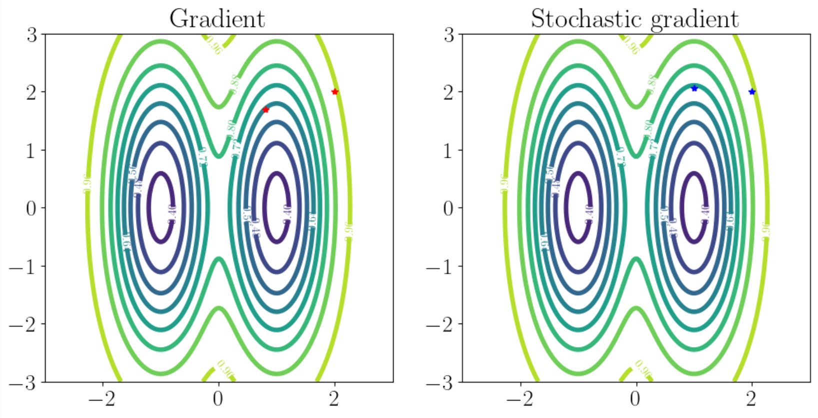



Gradient
Stochastic gradient
3
2
1
0
−1
−2
−3
−2
0
2

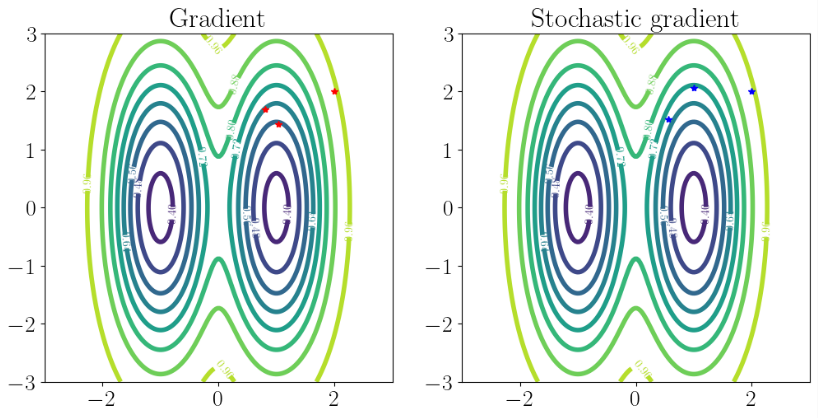
A starting problem



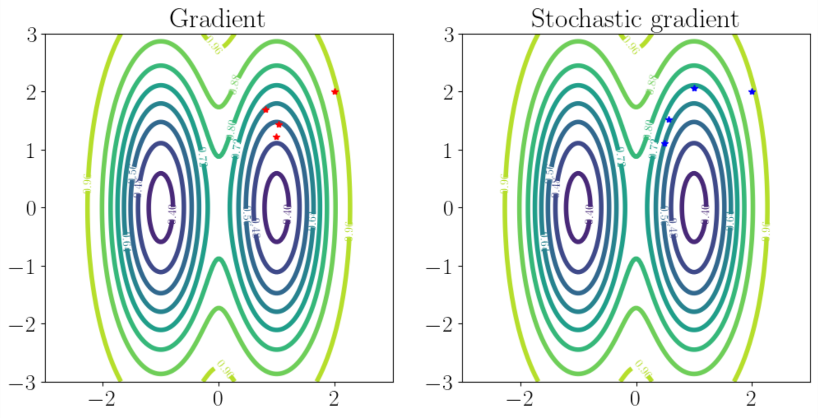
A starting problem



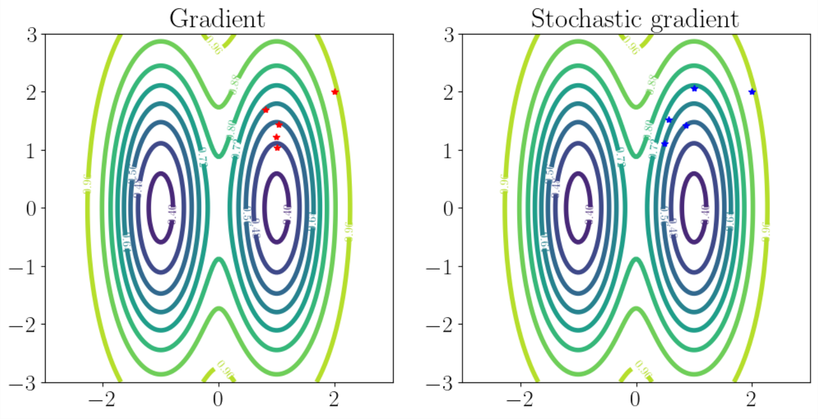
A starting problem



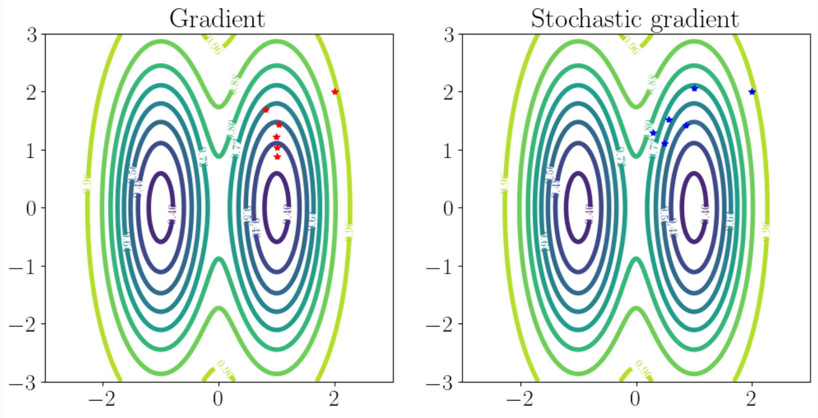
A starting problem



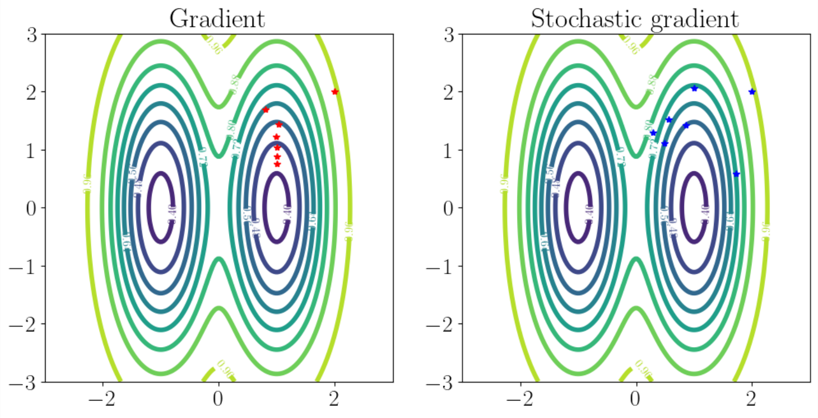
A starting problem



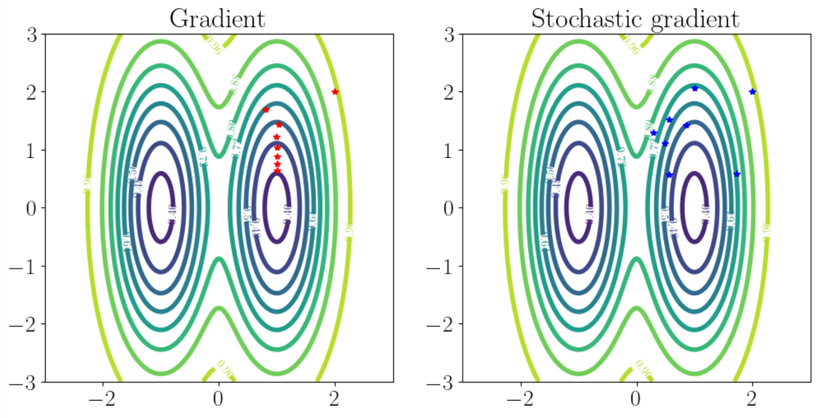
A starting problem



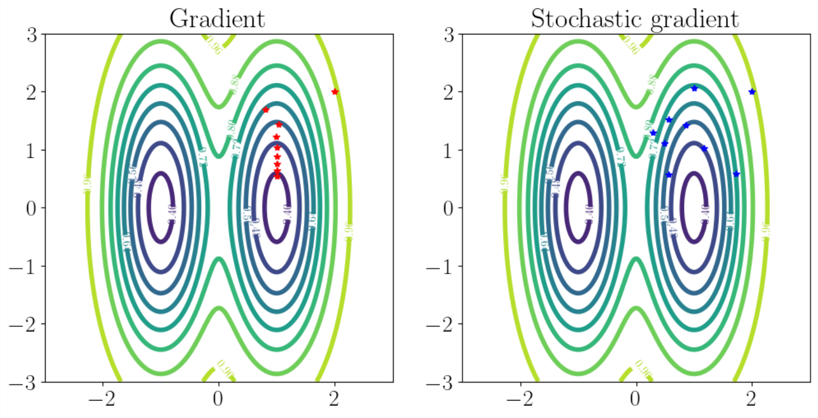
A starting problem



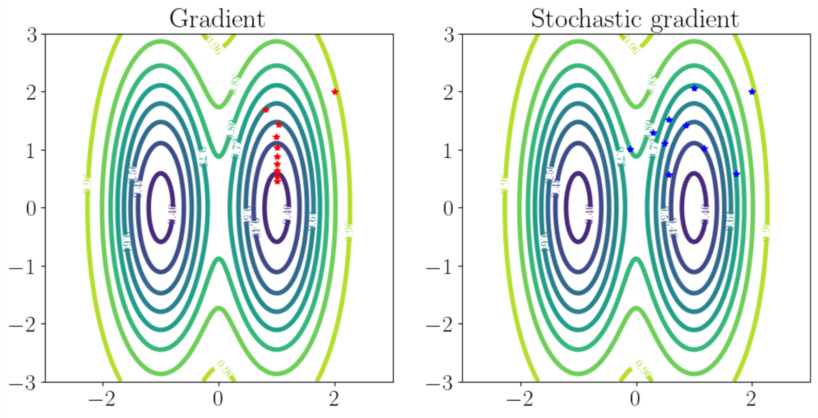
A starting problem



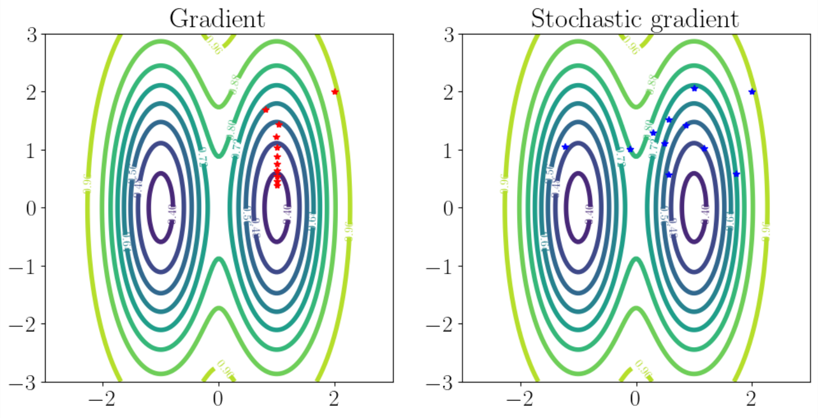
A starting problem



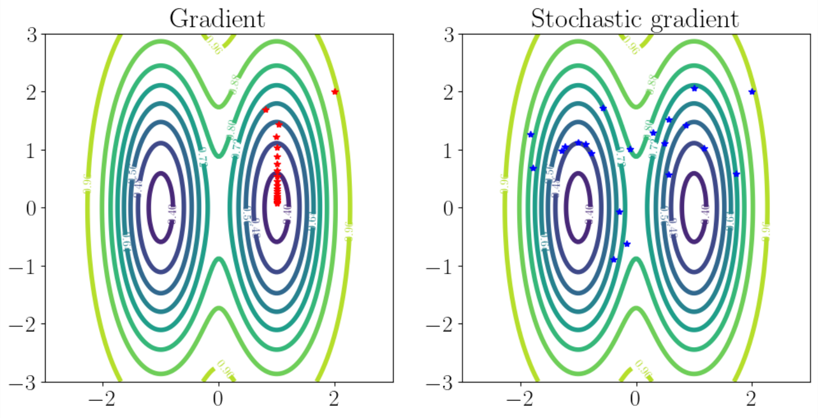
A starting problem



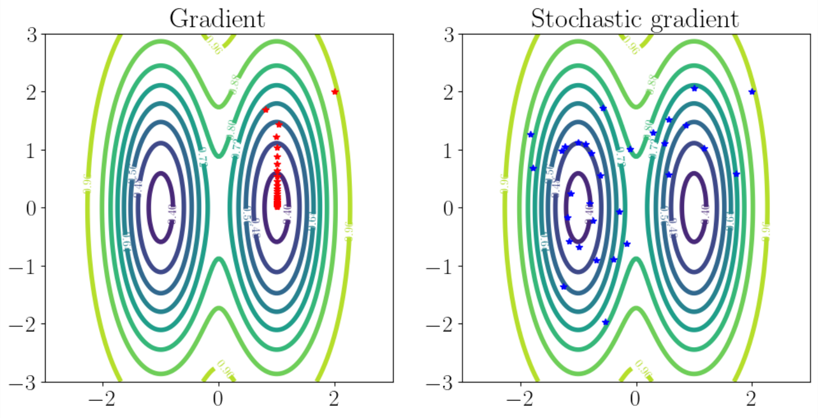
A starting problem



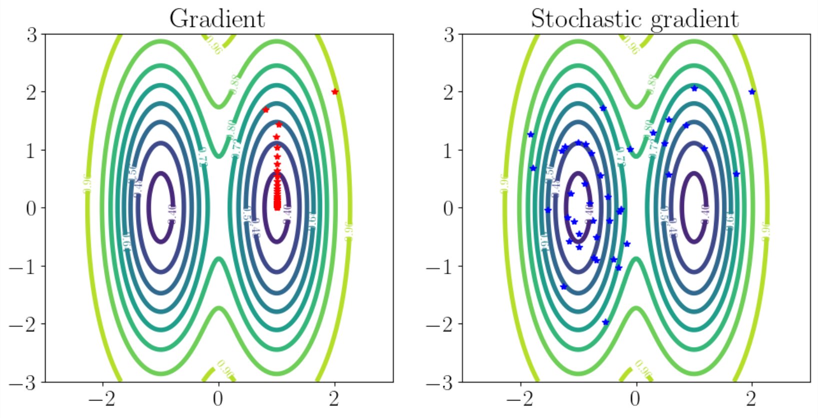
A starting problem



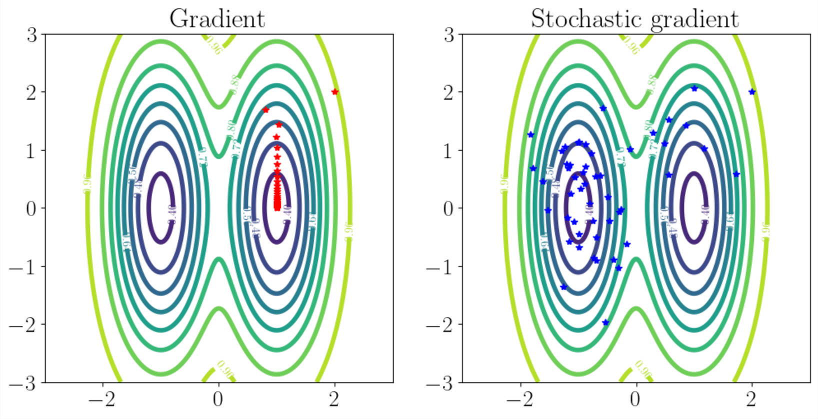
A starting problem



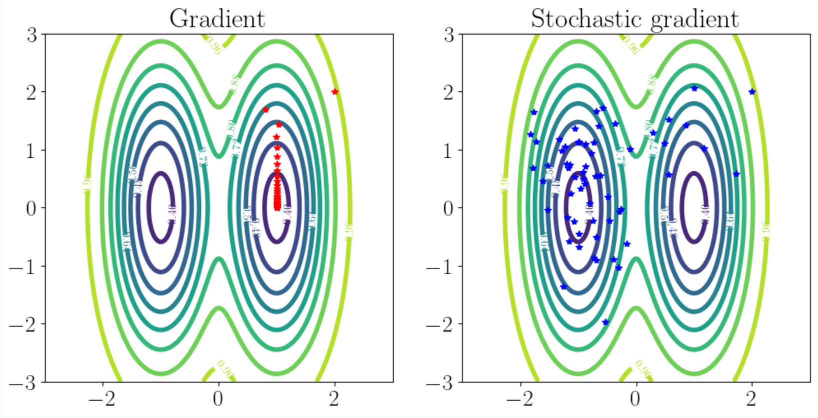
A starting problem



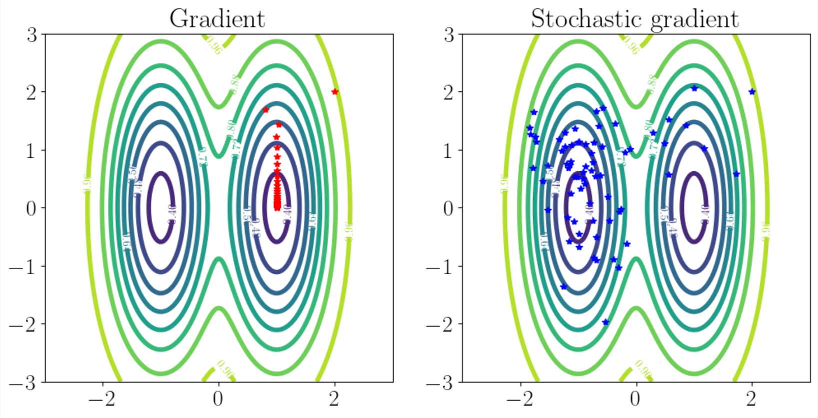
A starting problem



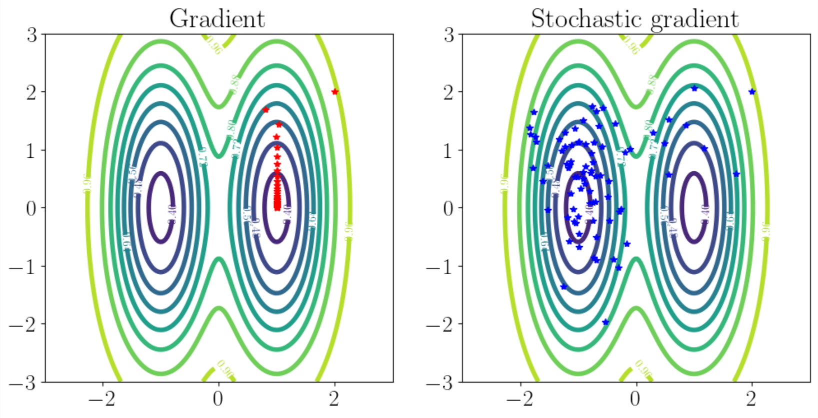
A starting problem



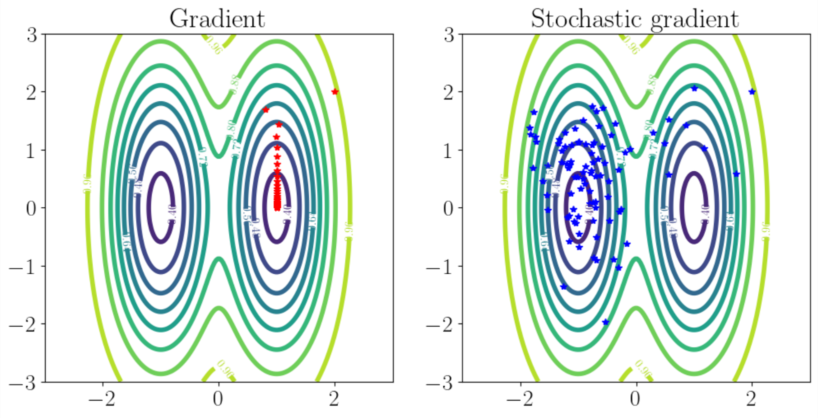
A starting problem



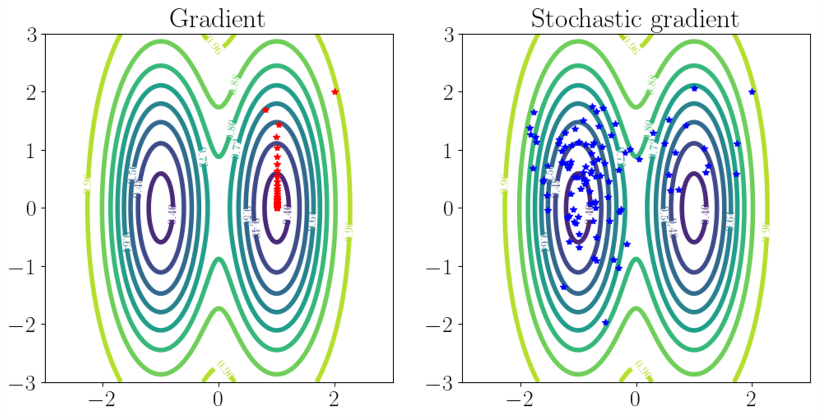
A starting problem



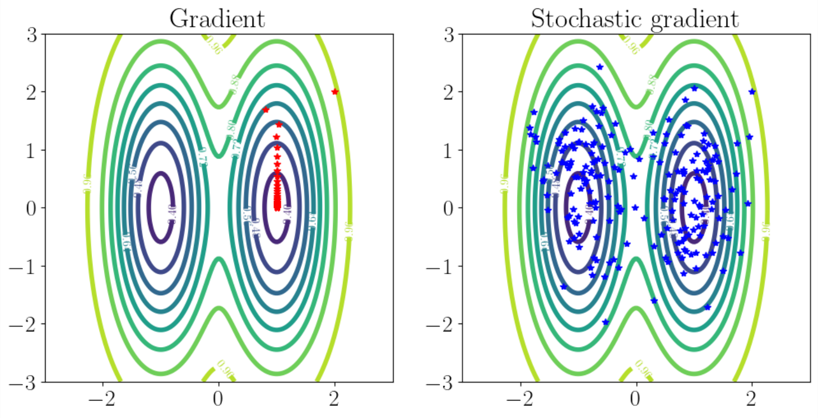
A starting problem



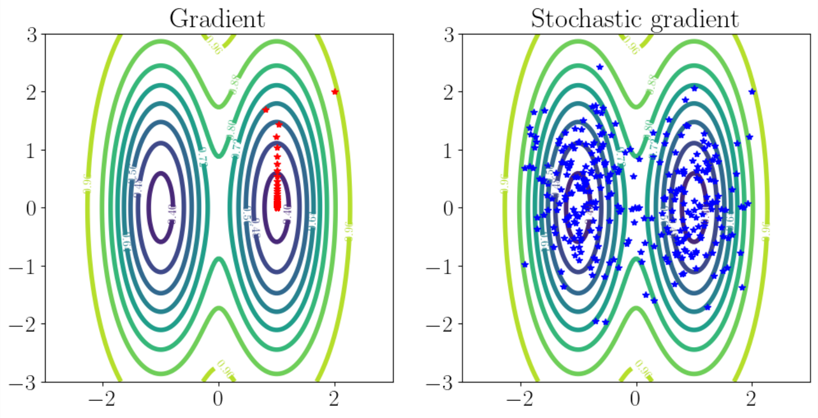
A starting problem



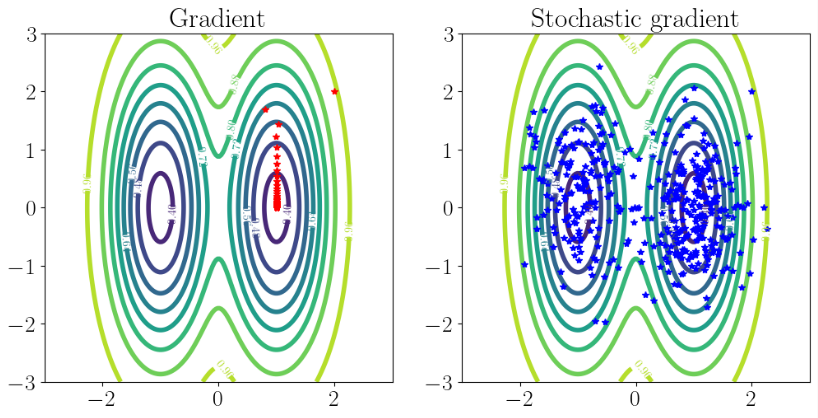
A starting problem

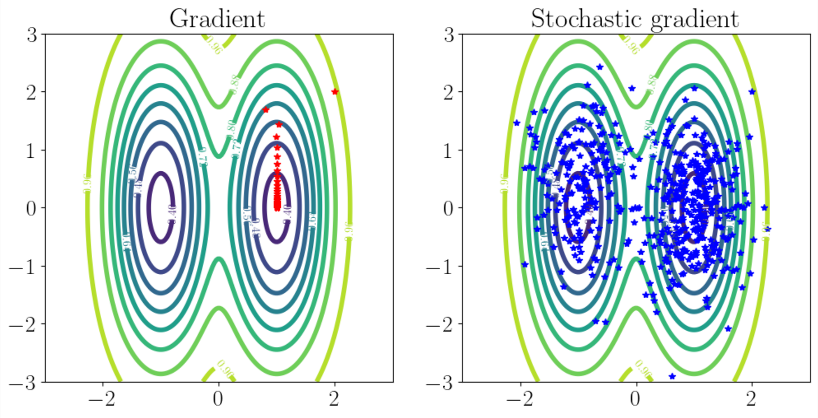


A starting problem

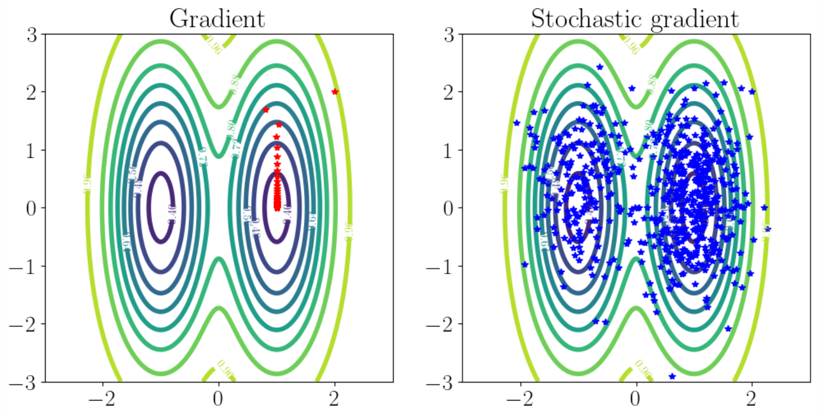


A starting problem

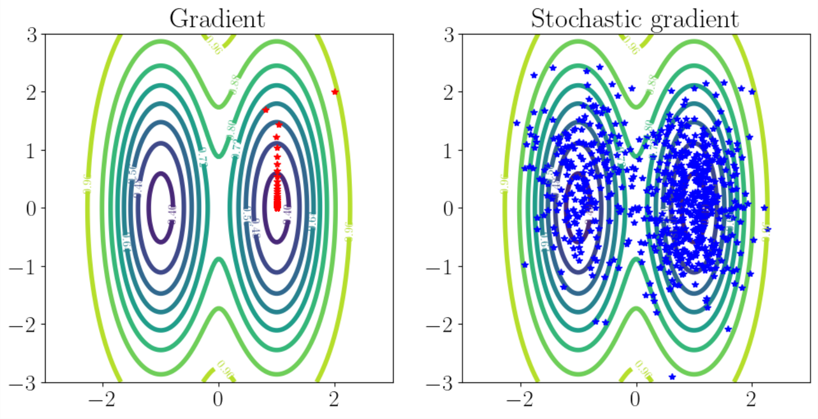



Gradient
Stochastic gradient
3
2
1
0
−1
−2
−3
−2
0
2

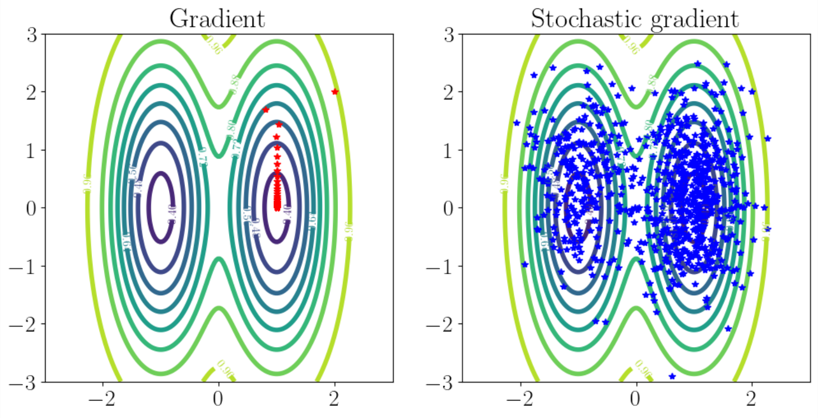
A starting problem



A starting problem



A starting problem



Computing the MMSE (Monte-Carlo – slow)

Run the Euler-Maruyama scheme and

$$\hat{x}_{MMSE}(y) = \mathbb{E}[X|Y = y] \approx \frac{1}{K}\sum_{k=1}^{K} x_k.$$

Computing the MMSE (Monte-Carlo – slow)

Run the Euler-Maruyama scheme and

$$\hat{x}_{MMSE}(y) = \mathbb{E}[X|Y = y] \approx \frac{1}{K}\sum_{k=1}^{K} x_k.$$

Computing the MMSE (Unrolled networks – fast)

Assume that $Y = A(X) + B$.

Construct a sequence of denoising networks $N_k(x, w_k)$, $k = 1 \ldots K$:

$$\begin{aligned} x_0 &= A^+ y \\ x_{k+\frac{1}{2}} &= x_k - \tau \nabla \log p_{Y|X}(y|x_k) \\ x_{k+1} &= x_k - \tau \frac{x_k - N(x_k, w_k)}{\delta^2} \end{aligned}$$

Define the architecture $\mathcal{UN}(y, w) = x_K$ with $w = (w_1, \ldots, w_K)$.

After training:

$$\mathcal{UN}(y, w^\star) \approx \hat{x}_{MMSE}(y)!$$

Main facts

- Learn to denoise!
  - MMSE denoising
  - $\approx$ prior via $\nabla \log p_X$ (Tweedie formula)
  - Universal: can be used for arbitrary inverse problems

Main facts

- Learn to denoise!
    - MMSE denoising
    - $\approx$ prior via $\nabla \log p_X$ (Tweedie formula)
    - Universal: can be used for arbitrary inverse problems
- MAP estimation
    - Plug&play (universal method)
    - “Best” looking result
    - Can be slow at runtime... But, can we trust it?

Main facts

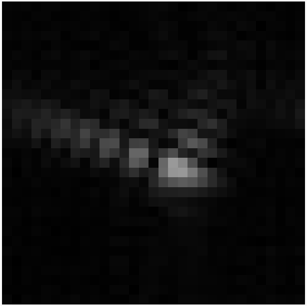
- Learn to denoise!
  - MMSE denoising
  - $\approx$ prior via $\nabla \log p_X$ (Tweedie formula)
  - Universal: can be used for arbitrary inverse problems
- MAP estimation
  - Plug&play (universal method)
  - "Best" looking result
  - Can be slow at runtime... But, can we trust it?
- Posterior sampling
  - Obtained by SDEs (diffusion models)
  - The Bayesian Graal!
  - Really slow
  - Heavy ongoing research

Main facts

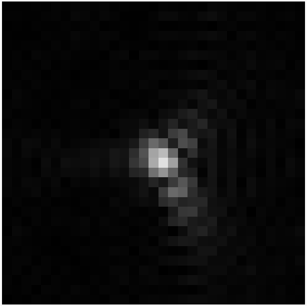
- Learn to denoise!
    - MMSE denoising
    - $\approx$ prior via $\nabla \log p_X$ (Tweedie formula)
    - Universal: can be used for arbitrary inverse problems
- MAP estimation
    - Plug&play (universal method)
    - "Best" looking result
    - Can be slow at runtime... But, can we trust it?
- Posterior sampling
    - Obtained by SDEs (diffusion models)
    - The Bayesian Graal!
    - Really slow
    - Heavy ongoing research
- MMSE
    - Unrolled network (specific to an operator)
    - "Best" result in average (blurry where unfaithful)
    - Fast at runtime
    - Long at train time

Mambo applications

Unrolled networks (fast MMSE) and adaptivity issues



Operator (PSF) A_0



Operator (PSF) A_1

Image deblurring

Unrolled networks (fast MMSE) and adaptivity issues

Ground truth x

Unrolled networks (fast MMSE) and adaptivity issues

Blurry image $y_0 = A_0(x) + b$, 19.1dB

Learning on operator families

Unrolled networks (fast MMSE) and adaptivity issues

No mismatch $\mathcal{N}(w_0^*, A_0, y_0)$ 30.1dB

Unrolled networks (fast MMSE) and adaptivity issues

Adaptivity issue $\mathcal{N}(w_0^*, A_1, y_1)$ 15.6dB

Unrolled networks (fast MMSE) and adaptivity issues

Model mismatch $\mathcal{N}(w_0^*, A_0, y_1)$ 14.5dB

Learning on operator families

Train on operator families!

$$\inf_{w} \mathbb{E}_{X,A,Y} \left[\| \mathcal{UN}(X, A, w) - Y \|_2^2 \right]$$

- Do not rely on the generalization capacity
- No performance loss
- Possibility to use in blind inverse problems

Blind deblurring & the MAP

Assume that $y = h \star x + b$

- h: unknown PSF
- x: unknown sharp image

Blind deblurring & the MAP

Assume that $y = h \star x + b$

- h: unknown PSF
- x: unknown sharp image

Claim

For a "natural image" prior p_X

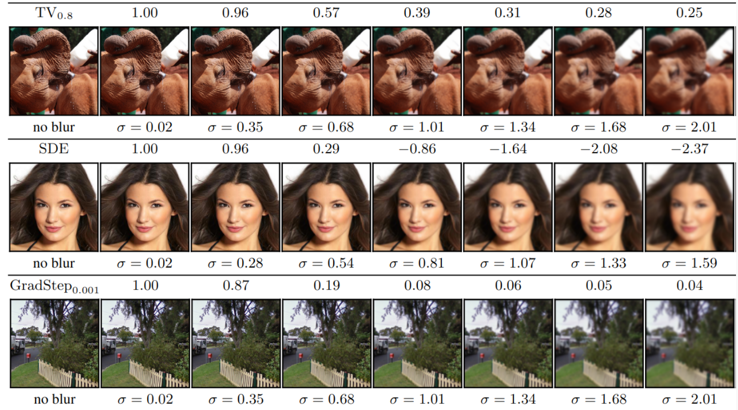
Blind deblurring & the MAP

Assume that $y = h \star x + b$

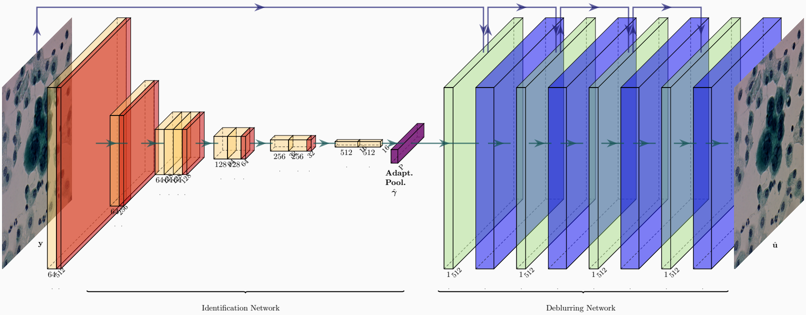
- h: unknown PSF
- x: unknown sharp image

Claim

For a "natural image" prior $p_X \Rightarrow$ MAP yields $h^\star = \delta$ and blurry solutions!



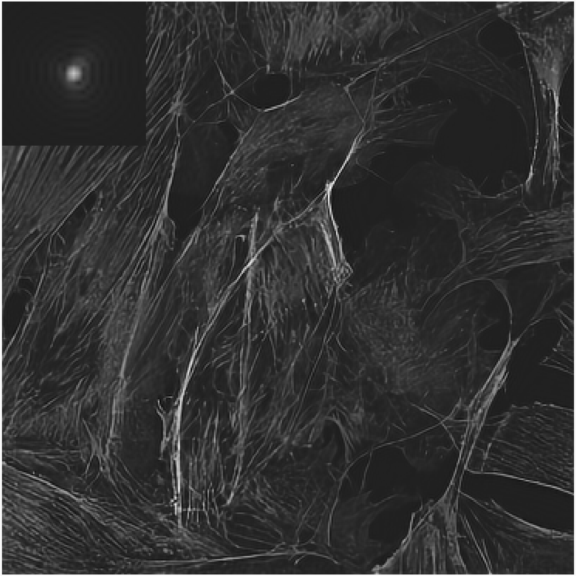
Blind deblurring & the MMSE



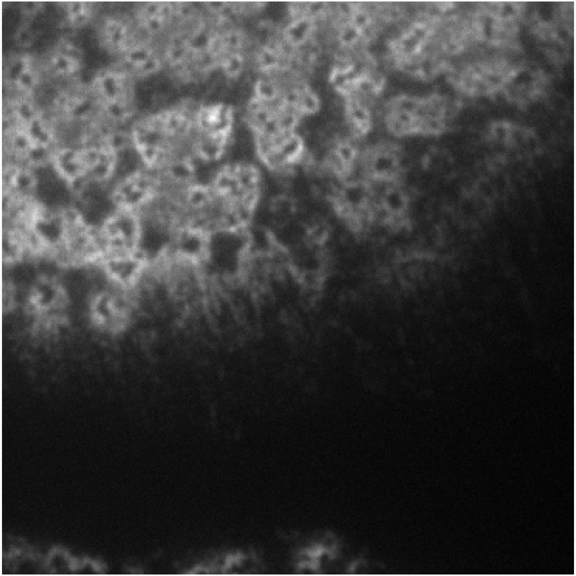
A specific architecture

Blind deblurring & the MMSE

Real image (actin filaments)

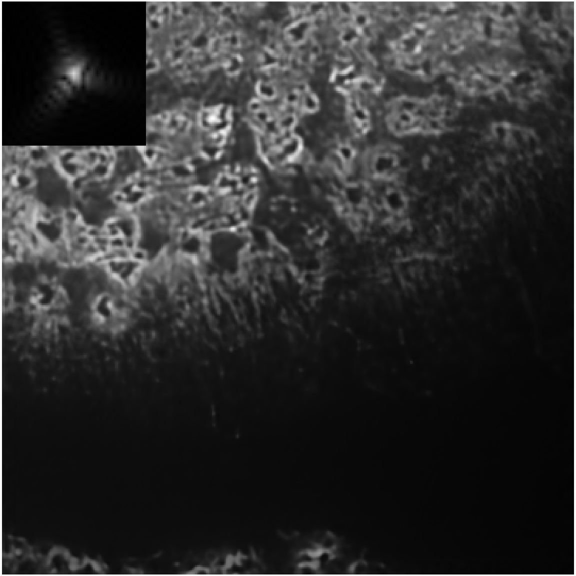


Deep-Blur result



Real image (R. Poincloux, podosomes)

Blind deblurring & the MMSE

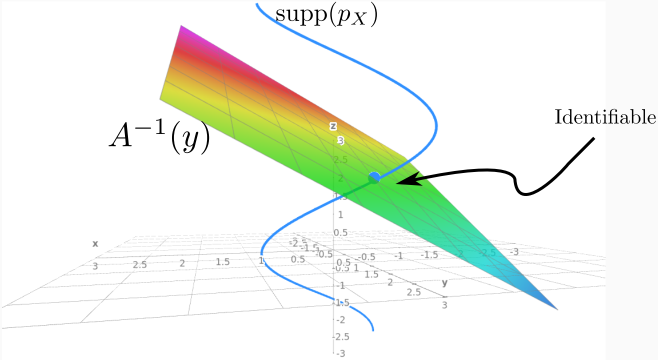


Deep-Blur results

Certifying blind inverse problems

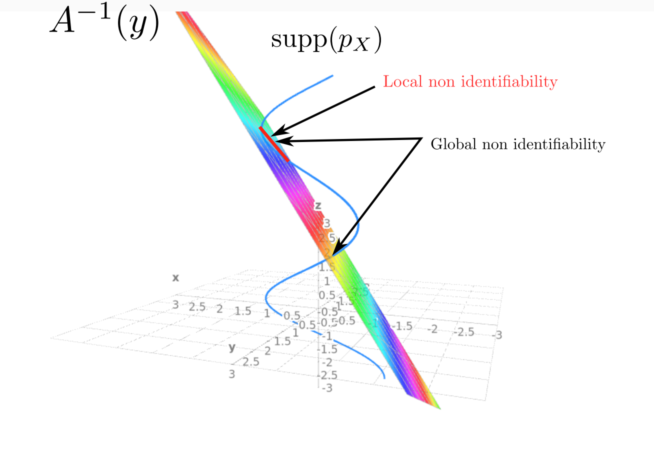
Some inverse problems are easier than others!

$\dim(\ker(A))$ doesn't matter!



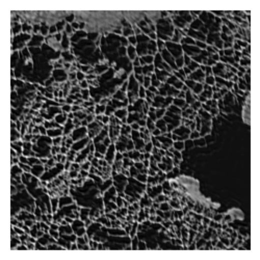
A favorable inverse problem

Certifying blind inverse problems

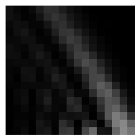


A problematic inverse problem

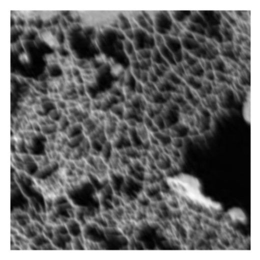
Certifying blind inverse problems



$\star$



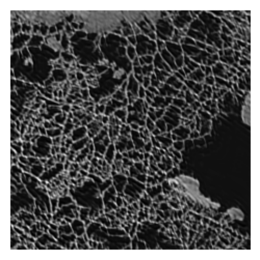
$=$



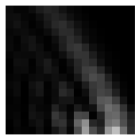
Estimator $\mathcal{D}(y, x)$ Kernel h_x $\Phi(x)$

Exploring equally likely images

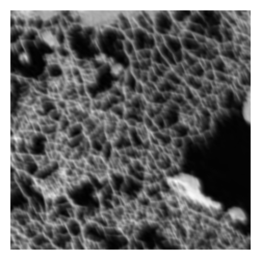
Certifying blind inverse problems



$\star$



$=$



Estimator $\mathcal{D}(y, x)$ Kernel h_x $\Phi(x)$

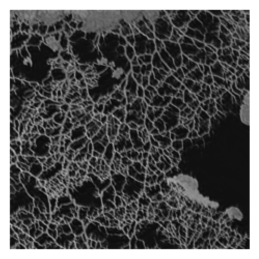
Exploring equally likely images

Certifying blind inverse problems

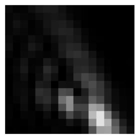
Estimator $\mathcal{D}(y, x)$ Kernel h_x $\Phi(x)$

Exploring equally likely images

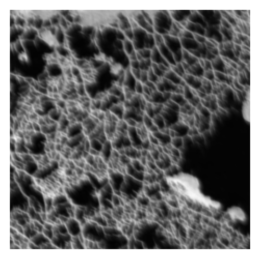
Certifying blind inverse problems



$\star$

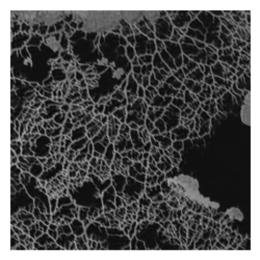


$=$

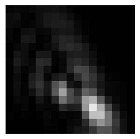


Estimator $\mathcal{D}(y, x)$ Kernel h_x $\Phi(x)$

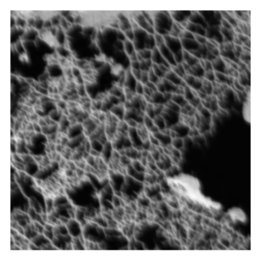
Exploring equally likely images



$\star$



$=$



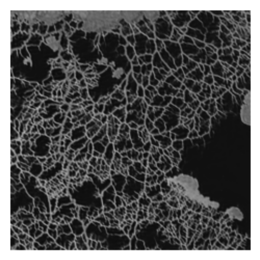
Estimator $\mathcal{D}(y, x)$

Kernel h_x

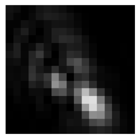
$\Phi(x)$

Exploring equally likely images

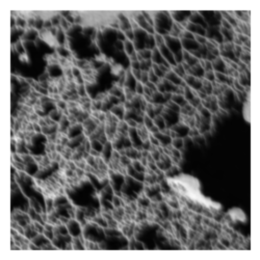
Certifying blind inverse problems



$\star$

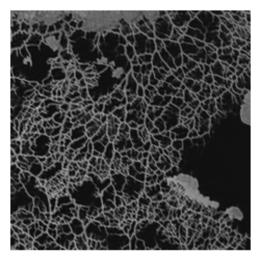


$=$

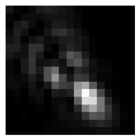


Estimator $\mathcal{D}(y, x)$ Kernel h_x $\Phi(x)$

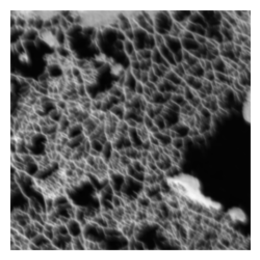
Exploring equally likely images



$\star$



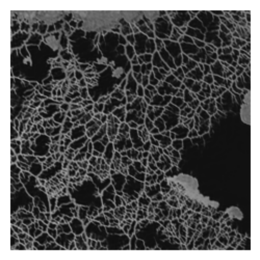
$=$



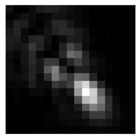
Estimator $\mathcal{D}(y, x)$ Kernel h_x $\Phi(x)$

Exploring equally likely images

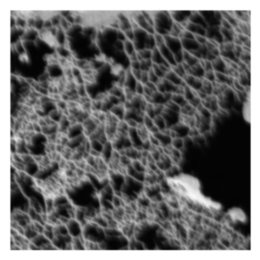
Certifying blind inverse problems



$\star$



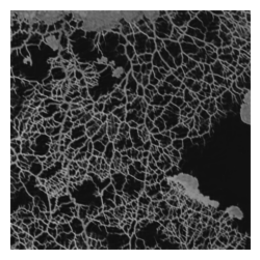
$=$



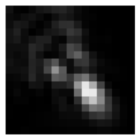
Estimator $\mathcal{D}(y, x)$ Kernel h_x $\Phi(x)$

Exploring equally likely images

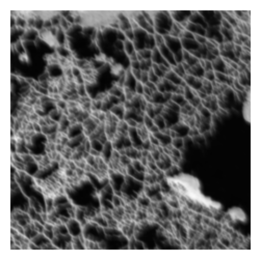
Certifying blind inverse problems



$\star$



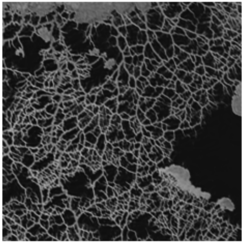
$=$



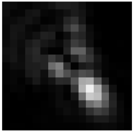
Estimator $\mathcal{D}(y, x)$ Kernel h_x $\Phi(x)$

Exploring equally likely images

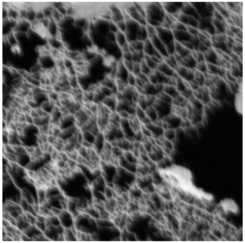
Certifying blind inverse problems



$\star$



$=$



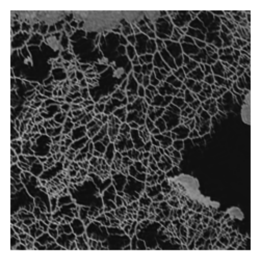
Estimator $\mathcal{D}(y, x)$

Kernel h_x

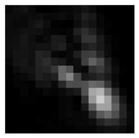
$\Phi(x)$

Exploring equally likely images

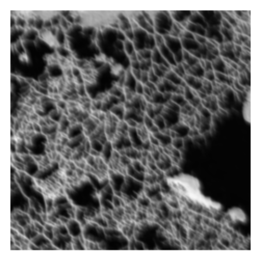
Certifying blind inverse problems



$\star$



$=$



Estimator $\mathcal{D}(y, x)$ | Kernel h_x | $\Phi(x)$

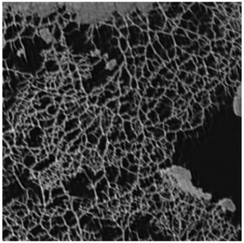
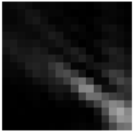
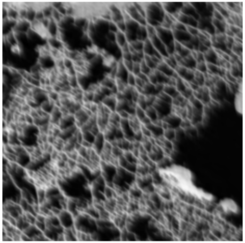
Exploring equally likely images

Certifying blind inverse problems

Estimator $\mathcal{D}(y, x)$ Kernel h_x $\Phi(x)$

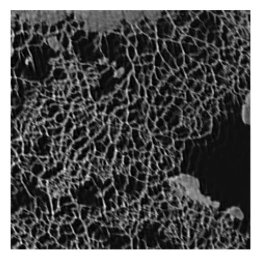
Exploring equally likely images

Certifying blind inverse problems

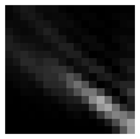
 $\star$  $=$ 

Estimator $\mathcal{D}(y, x)$ Kernel h_x $\Phi(x)$

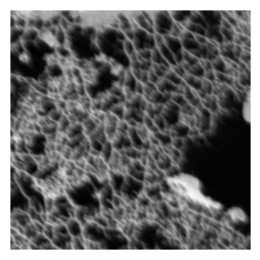
Exploring equally likely images



$\star$



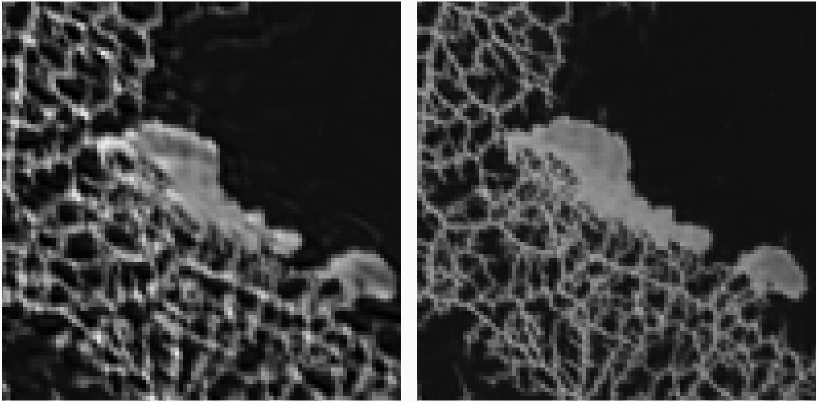
$=$



Estimator $\mathcal{D}(y, x)$ Kernel h_x $\Phi(x)$

Exploring equally likely images

Certifying blind inverse problems



Details of two possible deconvolutions

Optimizing Fourier sampling trajectories

The MRI model

$$y = A(\theta)x + b$$

- θ: sampling locations in the Fourier domain

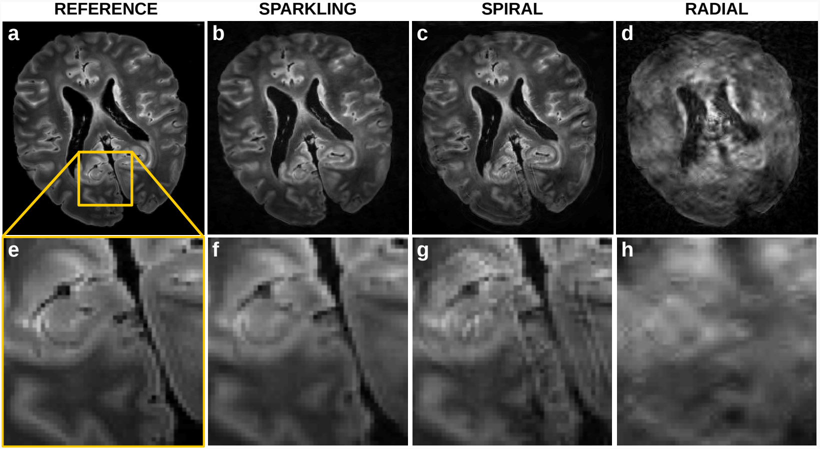
How to optimally sample?

Compressed sensing
DL Donoho - IEEE Transactions on information theory, 2006 - ieeexplore.ieee.org
Suppose x is an unknown vector in Ropf m (a digital image or signal); we plan to measure n general linear functionals of x and then reconstruct. If x is known to be compressible by …
☆ Enregistrer 99 Citer Cité 34069 fois Autres articles Les 25 versions Web of Science: 18790

Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information
EJ **Candès**, J Romberg, T **Tao** - IEEE Transactions on …, 2006 - ieeexplore.ieee.org
This paper considers the model problem of reconstructing an object from incomplete frequency samples. Consider a discrete-time signal f/spl isin/C/sup N/ and a randomly chosen set of …
☆ Enregistrer 99 Citer Cité 19447 fois Autres articles Les 29 versions Web of Science: 10638

Optimizing Fourier sampling trajectories

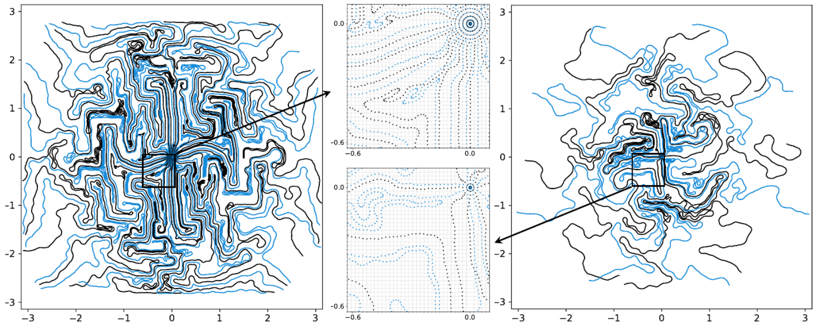
Optimizing Fourier sampling trajectories



Cartesian (2’20), SPARKLING (8.8"), Spiral (8.8"), Radial (8.8")

Joint sampling/reconstruction scheme optimization

$$\inf_{w,\theta} \mathbb{E}_{X,Y}\left[\|N(Y,A(\theta),w)-X\|_2^2\right]$$



(c) Joint traj. optim. 25%

(d) Joint traj. optim. 10%

Significantly higher performance

Conclusion

Bayes & Imaging

- Pre 2015 priors used to be too far from reality...

Conclusion

Bayes & Imaging

- Pre 2015 priors used to be too far from reality...
- Experiments suggest that learnt priors are now rocking!
- Provides an elegant mathematical framework

Conclusion

Bayes & Imaging

- Pre 2015 priors used to be too far from reality...
- Experiments suggest that learnt priors are now rocking!
- Provides an elegant mathematical framework

Perspectives

- How reliable is this interpretation?
  - Pope et al, The intrinsic dimension of images (is $\approx$ 40), *ICLR* 2021

Conclusion

Bayes & Imaging

- Pre 2015 priors used to be too far from reality...
- Experiments suggest that learnt priors are now rocking!
- Provides an elegant mathematical framework

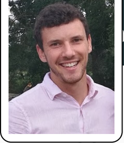
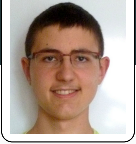
Perspectives

- How reliable is this interpretation?
    - Pope et al, The intrinsic dimension of images (is ≈ 40), *ICLR* 2021
    - 10^9 samples are not enough to learn an p.d.f. in dim 40!
- What neural architectures promote natural images? Are there things beyond CNNs? What are the biases?
- Why can we optimize/train near globally?
- How to accelerate computations to quantify uncertainty?
- Plenty new applications in imaging / system biology!

A few personal references

More details

- Gossard & P.W., Training adaptive reconstruction networks for blind inverse problems, *SIAM Imaging Science* 2024
- Debarnot & P.W., DEEP-BLUR: Blind Identification and Deblurring with CNN, *Biological Imaging*, 2024
- Munier, Soubies & P.W., Identifying the non identifiable, ongoing, 2024
- Nguyen, Pauwels & P.W., Don't use the MAP blindly, ongoing, 2024
- Lazarus, Ciuciu & P.W. SPARKLING: variable density filling curves for MRI, *Magnetic Resonance in Medicine*, 2019
- Gossard & P.W., Bayesian Optimization of Sampling Densities in MRI, *MELBA* 2023

More details

- Zhu et al, Denoising Diffusion Models for Plug-and-Play Image Restoration, *CVPR* 2023
- Laumont et al, Bayesian plug & play priors: when Langevin meets Tweedie, *SIAM Imaging Science*, 2022
- Vahdat et al, 4h Nvidia Tutorial on Denoising diffusion models, 2023


Main collaborators

- Valentin Debarnot (Swiss post-doc, who may apply to INRA)
- Alban Gossard (@Go Pro)
- Carole Lazarus (@Siemens research)
- Nathanaël Munier, (current PhD)
- Minh Hai Nguyen, (current PhD)
- F. de Gournay, P. Escande, E. Soubies, E. Pauwels, J. Kahn