Bayesian imaging using Plug & Play priors.

When Langevin meets Tweedie.

Rémi Laumont

Joint work with:
Valentin De Bortoli, Julie Delon, Andrés Almansa, Alain Durmus, Marcelo Pereyra.

June 24, 2021

Introduction to the image restoration context.

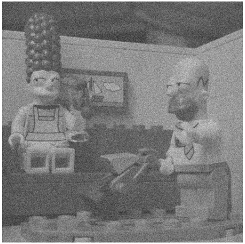
- Modelisation:

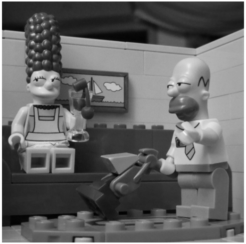
$$\boxed{y = A(x) + n}$$

with $x \in \mathbb{R}^d$ the unknown scene, $y \in \mathbb{R}^m$ the observation, $n \in \mathbb{R}^m$ the noise, and $A : \mathbb{R}^d \to \mathbb{R}^m$ a known degradation operator.

- Classical Goal: Estimate x from its observation y.

Examples.





Noisy observation y. True scene x. Blurry observation y.

Problem: ill-posed, -conditionned. $\rightarrow$ Need to regularize.

Bayesian paradigm.

- Bayesian formulation:

$$p(x|y) = \frac{p(x)p(y|x)}{\int_{\mathbb{R}^d} p(\tilde{x})p(y|\tilde{x})\mathrm{d}\tilde{x}} \propto p(x)p(y|x)$$

where $p(x)$ the prior and $p(y|x)$ is the likelihood (assumed to be known).

- $R(x) = -\log p(x)$ and $F(x, y) = -\log p(y|x)(= \frac{\|Ax-y\|_2^2}{2\sigma^2})$
- Maximum-A-Posteriori (MAP) estimator:

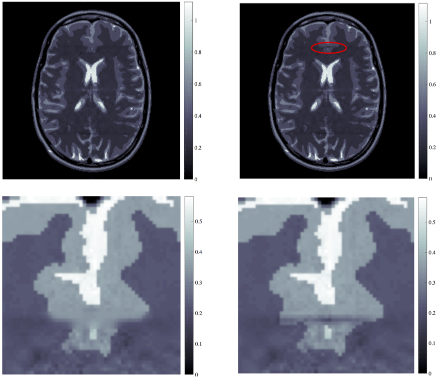
$$\boxed{\hat{x}_{MAP} = \underset{x\in\mathbb{R}^d}{\arg\max}\, p(x|y) = \underset{x\in\mathbb{R}^d}{\arg\min} \left\{F(x, y) + \lambda R(x)\right\}.} \quad (1)$$

- Minimum Mean Square Error (MMSE) estimator:

$$\boxed{\hat{x}_{MMSE} = \underset{u\in\mathbb{R}^d}{\arg\min}\, \mathbb{E}[\|x - u\|^2|y] = \mathbb{E}[x|y].} \quad (2)$$

Illustration of the limitations of the current restoration methods.

Brain restorations



Zooms

Why sampling from the posterior distribution?

Because it allows to:

- compute the MMSE estimation,
- perform uncertainty quantification,
- perform task such as model calibration.

Sampling using the Unadjusted Langevin Algortihm (ULA).

Goal: sampling from a distribution with target density $\pi(x) = p(x|y) \propto \exp(-R(x) - F(x, y))$.

- ULA:

$$X_{k+1} = X_k + \delta \nabla \log \pi(X_k) + \sqrt{2\delta} Z_{k+1}$$

$$\boxed{X_{k+1} = X_k - \delta \nabla R(X_k) - \delta \nabla F(X_k, y) + \sqrt{2\delta} Z_{k+1}} \quad (3)$$

  with $Z_k \sim \mathcal{N}(0, Id)$ for all $k \in \mathbb{N}$ and $\delta > 0$.

- Results:
  - Convergence towards a unique stationary distribution $\pi_\delta \neq \pi$ if δ small enough and $\nabla(R + F)$ is Lipschitz (Roberts, Tweedie, et al., 1996).
  - Convergence at exponential rate if $F + R$ is strongly convex at ∞ (Durmus and Moulines, 2017).

Plug-and-Play approaches.

- $p(x)$ (or $R(x)$) is unknown and difficult to model.
- Plug-and-Play methods aims at using a carefully chosen denosier $D_\epsilon : \mathbb{R}^d \to \mathbb{R}^d$ to implicitly define an image prior $p(x)$.
- Implicit prior to target ∇R e.g ((Alain and Bengio, 2014), (Guo et al., 2019),(Romano et al., 2017) and (Kadkhodaie and Simoncelli, 2020)) based on the Tweedie's formula.

Tweedie's formula

If $X \sim P_X$, $N \sim \mathcal{N}(0, Id)$ and $\tilde{X} = X + \sqrt{\epsilon}N$ then,

$$\boxed{\mathbb{E}[x|\tilde{x}] - \tilde{x} = \epsilon \nabla \log(p * g_\epsilon)(\tilde{x}) = \epsilon \nabla \log(p_\epsilon)(\tilde{x})},$$

with g_ϵ a Gaussian kernel with variance ϵ.

Plug-and-Play approaches and Tweedie's formula for sampling: PnP-ULA.

- Using a denoiser $D^*_\epsilon(\tilde{x}) = \mathbb{E}[x|\tilde{x}]$ we get

$$\boxed{D^*_\epsilon(\tilde{x}) - \tilde{x} = \epsilon \nabla \log(p_\epsilon)(\tilde{x})}.$$

- PnP-ULA:

$$X_{k+1} = X_k - \delta \nabla F(X_k, y) + \delta(D_\epsilon(X_k) - X_k)/\epsilon) \\ - \delta(\Pi_C(X_k) - X_k)/\lambda + \sqrt{2\delta} Z_{k+1}.$$

where this term ensures the strong convexity in the tails and Π_C is a projection on $B(0, R_C)$ and $D_\epsilon(x) \simeq D^*_\epsilon(x)$.

- Sampling from $\pi^C_{\delta,\epsilon}$.

Plug-and-Play approaches and Tweedie's formula for sampling: PnP-ULA (2).

- Hypotheses:
  - $Id - D_\epsilon$ is Lipschitz.
  - The likelihood $p(y|x)$ is bounded, C^1 and $\nabla \log p(y|x)$ is Lipschitz.
  - The MSE loss for D_ϵ^* is finite and uniformly bounded.
  - There exists $M : \mathbb{R}^+ \to \mathbb{R}^+$, such that for all $\|x\| \leq R$, $\|D_\epsilon(x) - D_\epsilon^*(x)\| \leq M(R)$.
- Convergence of π_ϵ towards π as $\epsilon \to 0$.
- Non-asymptotic error:

$$
\left|\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_k] - \int_{\mathbb{R}^d} \tilde{x} p(\tilde{x}|y) \mathrm{d}\tilde{x}\right| \leq C_0\{C_1 \epsilon^{\beta/4} + C_2 R_C^{-1} + C_3(\sqrt{\delta} + \frac{1}{n\delta} + C_R)\}. \quad (4)
$$

Problem position

- **Deblurring:** A encodes a bloc filter of size 9.
- **Inpainting:** A is a diagonal matrix with 1 or 0 on the diagonal and hidding 80% of the pixels in the original image.
- **Noise level:** $\sigma = 1/255$.
- **Original images:**

Simpson.

Cameraman.

Traffic.

Algorithm parameters

- D_ϵ provided by (Ryu et al., 2019) and such that $(D_\epsilon - Id)$ is L-Lipschitz with $L < 1$. $\epsilon = (5/255)^2$.
- Comparison with PnP-ADMM with $\epsilon_{deblurring} = (5/255)^2$ and $\epsilon_{inpainting} = (40/255)^2$.
- $C = [-1, 2]^d$.
- A thinned version of the Markov chain is considered made of samples stored every 2500 iterations.

	n	$n_{burn-in}$	δ	Initialization
PnP-ULA	2.5e7	2.5e6	$3\delta_{th}$	y

Deblurring results for Simpson.

Blurry image.

PSNR=22.44.

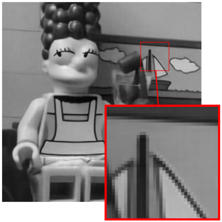
PnP-ULA.

PSNR=34.24.
SSIM=0.94.

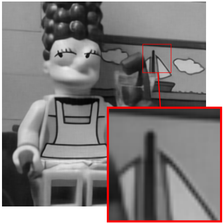
PnP-ADMM.

PSNR=32.48.
SSIM=0.93.

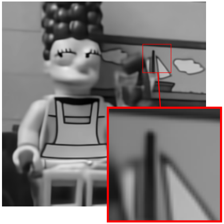
Detailed comparison between PnP-ULA and PnP-ADMM.



Original image.



PnP-ULA.



PnP-ADMM.

Deblurring results for Cameraman.

Blurry image.

PSNR=20.30.

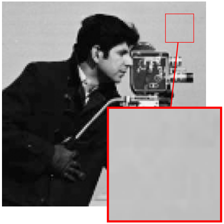
PnP-ULA.

PSNR=30.37.
SSIM=0.93.

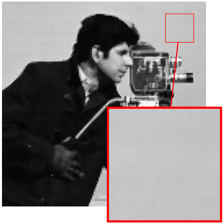
PnP-ADMM.

PSNR=30.81.
SSIM=0.89.

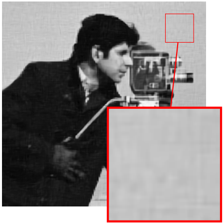
Detailed comparison between PnP-ULA and PnP-ADMM.



Original image.



PnP-ULA.

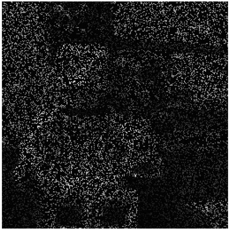


PnP-ADMM.

Deblurring results for traffic.

Blurry image	PnP-ULA	PnP-ADMM
		
PSNR=20.34.	**PSNR=29.86**.	PSNR=29.44.
	SSIM=0.89.	SSIM=0.87.

Inpainting results for Simpson.

Image to inpaint.	PnP-ULA.	PnP-ADMM.
		
PSNR=7.45.	**PSNR=31.51**.	PSNR=30.06.
	SSIM=0.94.	SSIM=0.92.

Inpainting results for Cameraman.

Image to inpaint.

PSNR=6.67.

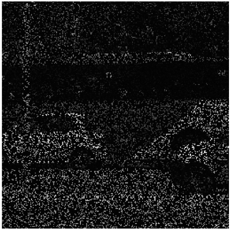
PnP-ULA.

PSNR=25.77.
SSIM=0.90.

PnP-ADMM.

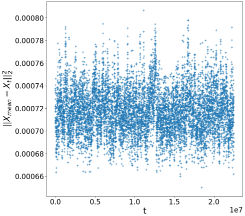
PSNR=24.80.
SSIM=0.90.

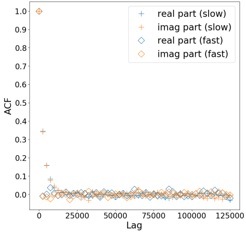
Inpainting results for traffic.

Image to inpaint.	PnP-ULA.	PnP-ADMM.
		
PSNR=8.35.	**PSNR=27.02**. **SSIM=0.85**.	PSNR=26.46. SSIM=0.84.

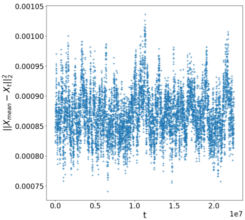
Convergence diagnosis for PnP-ULA

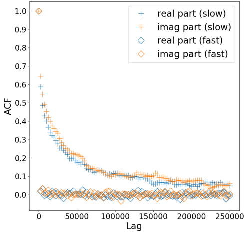
Deblurring.





Inpainting.



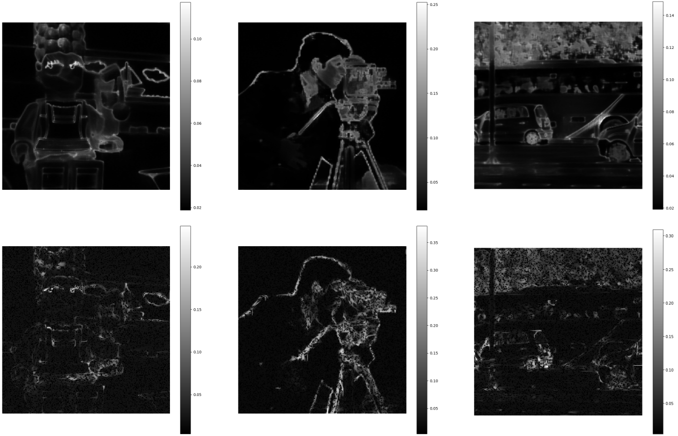


$\|MMSE - X_k\|_2^2$.

ACF.

Standard deviation estimates.

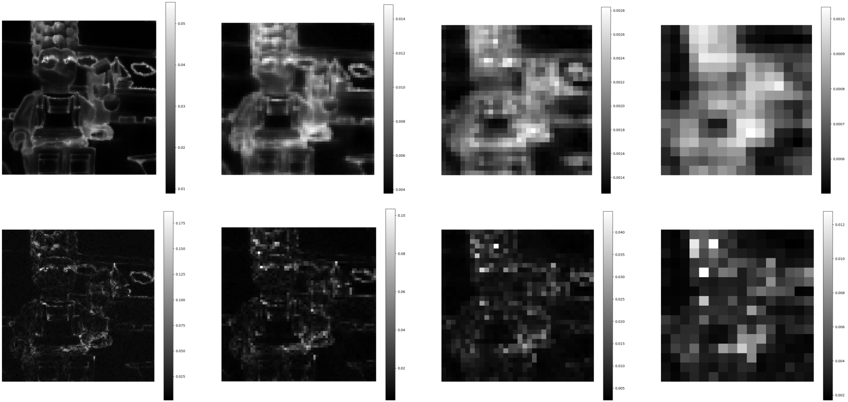
Deblurring.



Inpainting.

Standard deviation at different scales for Simpson.

Deblurring.



Inpainting.

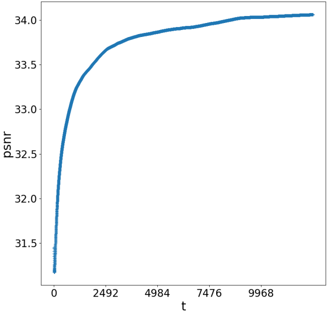
First conclusions.

- ✓ Computation of higher moments.
- ✓ First uncertainty study.
- ✓ Convergence Diagnosis.
- ✗ It is time-and memory-consuming ($\simeq$ 65 computation hours on Titan XP).
  $\rightarrow$ it depends on our goals.

Fast estimation of the MMSE for the deblurring problem.

Setting $\delta = 24\delta_{th}$, $n_{burn-in} = 0$ and TVL2 initialization.

MMSE estimate (PSNR $= 34.06$).



Evolution of the PSNR.

Convergence in approximately 2 minutes and 10^4 iterations!

Conclusion.

- Summary.
  - Development of a Langevin based algorithm with detailed convergence guarantees under realistic hypothesis.
  - Estimation of the non-asymptotic error when using this algorithm.
  - Efficient methods on different classical inverse problems.
- Future work.
  - Explore the performance of PnP-ULA with other MMSE denoisers having Lipschitz residual (*e.g.* NLMeans, SALSA).
  - Develop accelerated computation algorithms using this framework. (Pereyra et al., 2020).
  - Apply this framework to uncertainty quantification in astrophysics or medical imaging.

If you want to know more you can:

1. ask questions.
2. find the article related to this work on Arxiv
   https://arxiv.org/abs/2103.04715.

Alain, Guillaume and Yoshua Bengio (2014). "What Regularized Auto-Encoders Learn from the Data-Generating Distribution". In: *Journal of Machine Learning Research* 15, pp. 3743–3773. ISSN: 1532-4435. arXiv: 1211.4246 (cit. on p. 9).

Durmus, Alain and Éric Moulines (2017). "Nonasymptotic convergence analysis for the unadjusted Langevin algorithm". In: *Ann. Appl. Probab.* 27.3, pp. 1551–1587. ISSN: 1050-5164. DOI: 10.1214/16-AAP1238 (cit. on p. 8).

Guo, Bichuan, Yuxing Han, and Jiangtao Wen (2019). "AGEM: Solving Linear Inverse Problems via Deep Priors and Sampling". In: *Advances in Neural Information Processing Systems*, pp. 547–558 (cit. on p. 9).

Kadkhodaie, Zahra and Eero P Simoncelli (2020). "Solving Linear Inverse Problems Using the Prior Implicit in a Denoiser". In: *arXiv preprint arXiv:2007.13640* (cit. on p. 9).

Pereyra, Marcelo, Luis Vargas Mieles, and Konstantinos C. Zygalakis (2020). "Accelerating proximal Markov chain Monte Carlo by using an explicit stabilized method". In: *SIAM J. Imaging Sci.* 13.2, pp. 905–935. DOI: 10.1137/19M1283719 (cit. on p. 27).

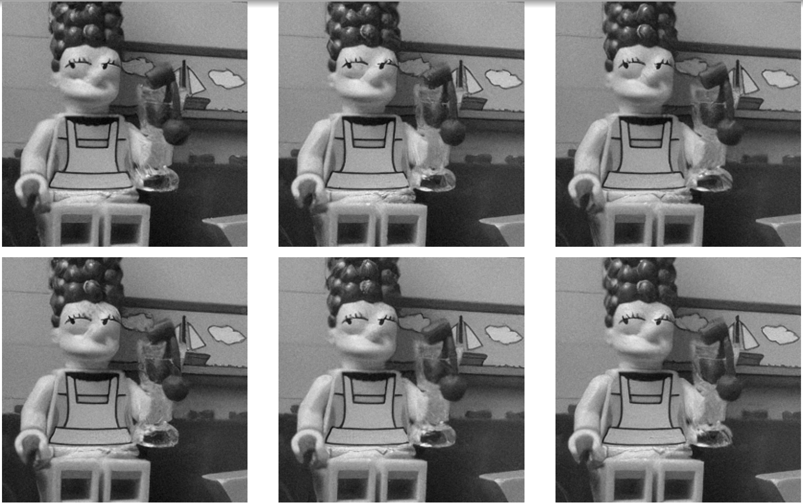
Roberts, Gareth O, Richard L Tweedie, et al. (1996). "Exponential convergence of Langevin distributions and their discrete approximations". In: *Bernoulli* 2.4, pp. 341–363 (cit. on p. 8).

Romano, Yaniv, Michael Elad, and Peyman Milanfar (2017). "The little engine that could: Regularization by denoising (RED)". In: *SIAM Journal on Imaging Sciences* 10.4, pp. 1804–1844 (cit. on p. 9).

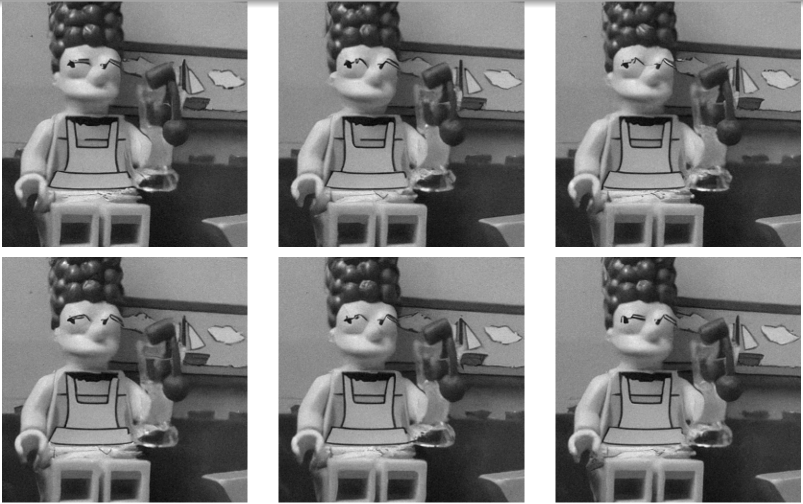
Ryu, Ernest K., Jialin Liu, Sicheng Wang, Xiaohan Chen, Zhangyang Wang, and Wotao Yin (2019). "Plug-and-Play Methods Provably Converge with Properly Trained Denoisers". In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, pp. 5546–5557. arXiv: 1905.05406 (cit. on p. 13).

Yan, Ke, Xiaosong Wang, Le Lu, and Ronald M Summers (2018). "DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning". In: *Journal of Medical Imaging* 5.3, p. 036501 (cit. on p. 34).

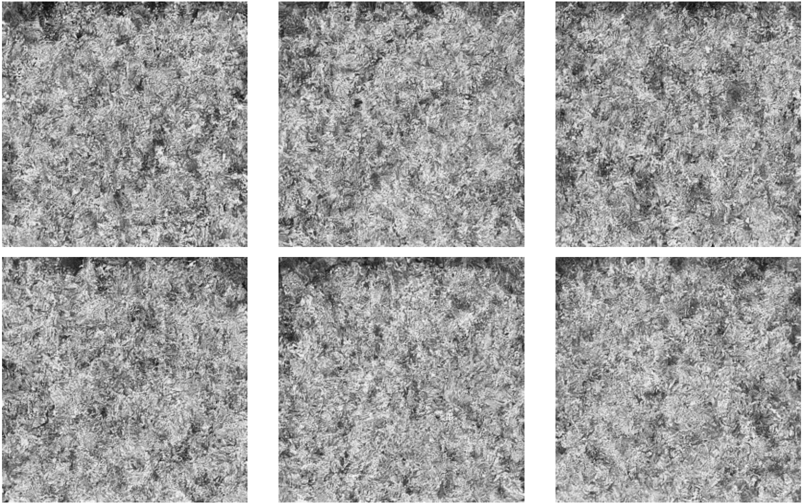
Samples of the posterior distribution for the deblurring problem.



Samples of the posterior distribution for the inpainting problem.

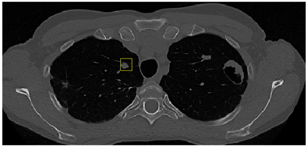


Samples of the prior distribution.



Problem position and parameter presentation.

- The original image x is a Computed Tomography slice from (Yan et al., 2018) with an annotated lesion.
- **Noise level:** $\sigma = 5/255$.
- We set $n = 5 \times 10^6$ with $n_{burn-in} = 5 \times 10^4$, $init = y$.
- We consider a thinned Markov chain made of samples stored every 500 iterations.



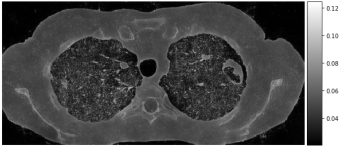
Original image from (Yan et al., 2018).

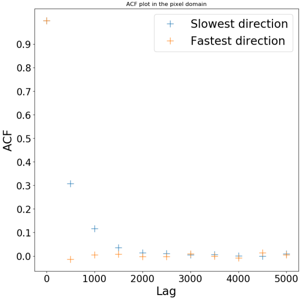
Noisy image (PSNR $= 20.17$).

First results.

MMSE estimate (PSNR = 29.12).

Standard deviation estimate.





ACF.

Uncertainty quantification on the lesion's size.

Among all samples stored we perform a simple segmentation to evaluate the lesion's size and plot the following histogram.