Bayesian imaging using Plug & Play priors.

When Langevin meets Tweedie.

Rémi Laumont

LABORATOIRE MAP5 · MATH INNOV · Région îledeFrance · Université de Paris

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

IUQ Workshop: September 27, 2022

1. Introduction.

2. Plug-and-Play priors for Bayesian imaging.
   - Bayesian framework.
   - Sampling using Langevin based methods.

3. Study of data-driven Plug-&-Play priors for sampling.
   - Visual Comparison between the SN-DnCNN and FINE priors.
   - Visual analysis of the Prox-GSD prior.
   - Potential analysis.
   - Frequentist accuracy of the Bayesian models.

4. Conclusion.

Introduction to the image restoration context.

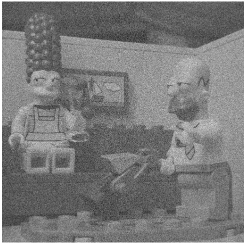
- Modelisation:

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

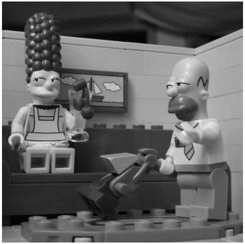
with $x \in \mathbb{R}^d$ the unknown scene, $y \in \mathbb{R}^m$ the observation, $n \in \mathbb{R}^m$ the noise, and $A : \mathbb{R}^d \rightarrow \mathbb{R}^m$ a known degradation operator.

- 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:

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

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 (Durmus and Moulines, 2017):
  - Convergence towards a unique stationary distribution $\pi_\delta \approx \pi$ if $\nabla(R+F)$ is L-Lipschitz and $\delta < 1/L$.
  - Exponentially fast convergence if $F + R$ is strongly convex at ∞ .

Plug-and-Play (PnP) approaches.

Problem: $p(x)$ (or $R(x)$) is unknown and difficult to model.

- PnP methods uses a denosier $D_\varepsilon : \mathbb{R}^d \to \mathbb{R}^d$ to implicitly define an image prior $p(x)$.
- Target ∇R ((Alain and Bengio, 2014), (Guo et al., 2019),(Romano et al., 2017) and (Kadkhodaie and Simoncelli, 2020)) using the Tweedie's formula.

Tweedie's formula

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

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

with g_ε a Gaussian kernel with variance ε.

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

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

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

- PnP-ULA:

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

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

- PPnP-ULA:

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

- Sampling respectively from $\pi^C_{\delta,\varepsilon}$ and $\pi^{C,P}_{\delta,\varepsilon}$.

Problem parameters

This talk: Investigating the influence of the implicit data-driven prior on the results.

- **Deblurring inverse problem**

$$y = Ax + n \tag{4}$$

  where A encodes a 9×9 bloc filter and $n \sim \mathcal{N}(0, \sigma^2)$.

- **Algorithms parameters**
  - $C = [-1, 2]^d$.
  - Initialization at the observation y.
  - $\delta = \delta_{th}$

PnP denoisers

- SN-DnCNN (Ryu et al., 2019)
  - CNN based architecture with ReLU activations.
  - $(D_\varepsilon - Id)$ is L-Lipschitz with $L < 1$.
  - $\varepsilon = (5/255)^2$.
  - Input: noisy grayscale image.
- FINE (Pesquet et al., 2020)
  - CNN based architecture with Leaky ReLU activations.
  - D_ε is firmly non-expansive $\Rightarrow$ 1-Lipschitz.
  - $\varepsilon = (2.25/255)^2$.
  - Input: noisy grayscale image.
- Prox-GSD (Hurault et al., 2022)
  - $D_\varepsilon = Id - \nabla g_\varepsilon$ with $g_\varepsilon(x) = \|x - N(x,\varepsilon)\|_2^2/2$ and $(D_\varepsilon - Id)$ 1-Lipschitz.
  - $N(.,\varepsilon)$: DRUNet based architecture with Softplus activations.
  - $\varepsilon = (15/255)^2$.
  - Inputs: noisy RGB image and noise map.

Comparison between the MMSE restorations of the SN-DnCNN and FINE induced priors within PnP-ULA for `Simpson`.

`Simpson`

Observation

PSNR=22.44/SSIM=0.66

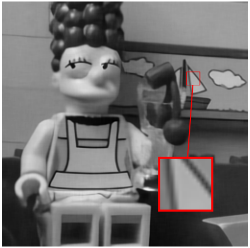
SN-DnCNN MMSE

PSNR=34.26/SSIM=0.94

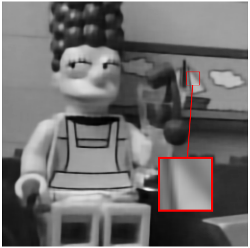
FINE MMSE

PSNR=31.77/SSIM=0.92

Detailed comparison between the MMSE restorations of the SN-DnCNN and FINE induced priors.



Sn-DnCNN MMSE



FINE MMSE

Comparison between the MMSE restorations of the SN-DnCNN and FINE induced priors within PnP-ULA for `Goldhill`.

`Goldhill`

Observation

PSNR=22.61/SSIM=0.45

SN-DnCNN MMSE

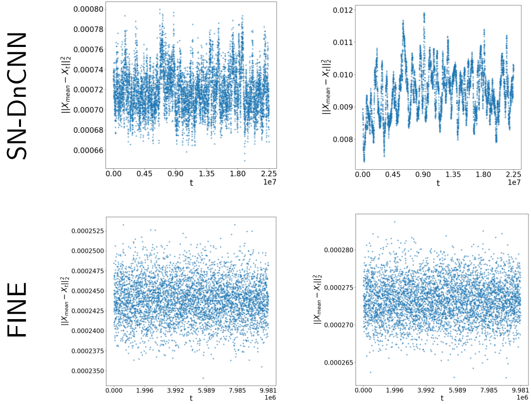
PSNR=27.72/SSIM=0.73

FINE MMSE

PSNR=27.20/SSIM=0.71

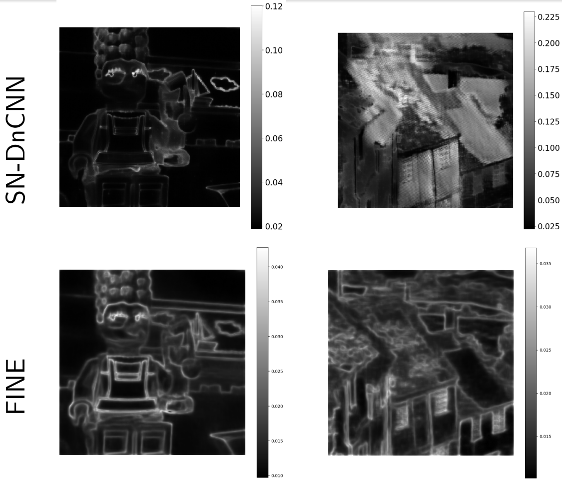
Zoom on the MMSE restoration produced by PnP-ULA with the SN-DnCNN induced prior for `Goldhill`.

L_2 distance between the final MMSE estimate and the samples generated by PnP-ULA with the SN-DnCNN and the FINE induced priors.

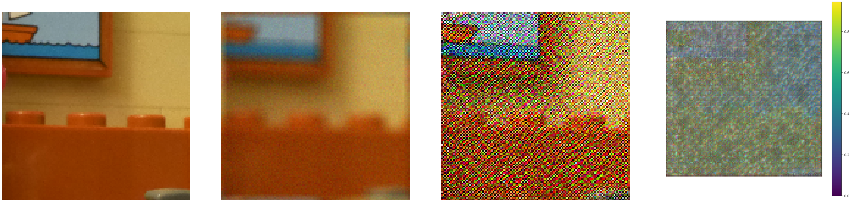


Simpson Goldhill

Comparison between the standard deviations produced of PnP-ULA with the SN-DnCNN and FINE induced priors.



Results for `Color Simpson` with PnP-ULA for the Prox-GSD prior.

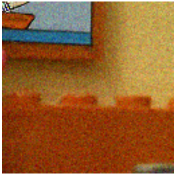


`Color Simpson` | Blurry observation PSNR=25.18 | MMSE | Standard deviation

Samples generated by PnP-ULA the Prox-GSD prior.

$n = 5e5$

$U(X_n) = 24112.473$



$n = 1.63e6$

$U(X_n) = 24089.926$

$n = 1.634e6$

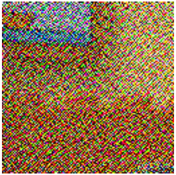
$U(X_n) = 24094.4$

$n = 1.64e6$

$U(X_n) = 24013.79$

$n = 1.7e6$

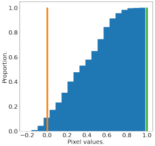
$U(X_n) = 24101.984$



$n = 5e6$

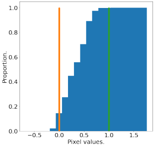
$U(X_n) = 22920.855$

Sample cumulative histograms.



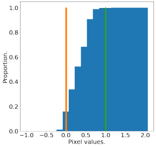
$n = 5e5$

$U(X_n) = 24112.473$



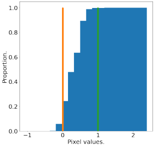
$n = 1.63e6$

$U(X_n) = 24089.926$



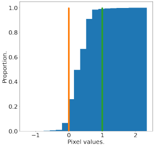
$n = 1.634e6$

$U(X_n) = 24094.4$



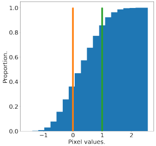
$n = 1.64e6$

$U(X_n) = 24013.79$



$n = 1.7e6$

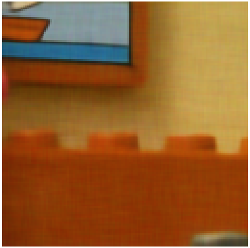
$U(X_n) = 24101.984$



$n = 5e6$

$U(X_n) = 22920.855$

Results obtained with PPnP-ULA and the Prox-GSD prior for `Color Simpson` ($C = [0,1]^d$).



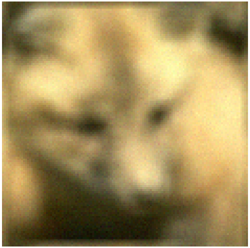
MMSE, PSNR=33.51

$n = 1e6$, PSNR=24.25

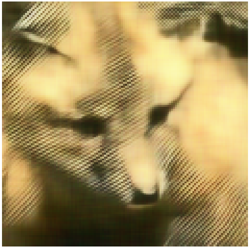
$n = 4e6$, PSNR=24.13

Results obtained with PPnP-ULA and the Prox-GSD prior for `Fox` ($C = [0,1]^d$).

`Fox`



Observation, PSNR=21.69



MMSE

Potential.

- Goal: Better understand the images promoted by our posterior.
- SN-DnCNN and FINE posterior potential: We consider the approximated potential U_{red} introduced in (Romano et al., 2017):

$$\forall x \in \mathbb{R}^d, \; U_{red}(x) = \frac{\|Ax - y\|^2}{2\sigma^2} + \frac{\alpha}{2} x^T (x - D_\varepsilon(x)). \tag{5}$$

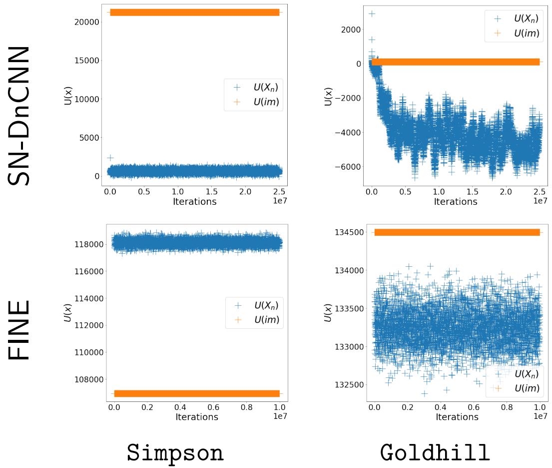
with $\alpha = 1/\varepsilon$ and such that

$$\nabla U_{red}(x) = -\frac{A^T(Ax - y)}{\sigma^2} + \frac{1}{\varepsilon}(x - D_\varepsilon(x)).$$

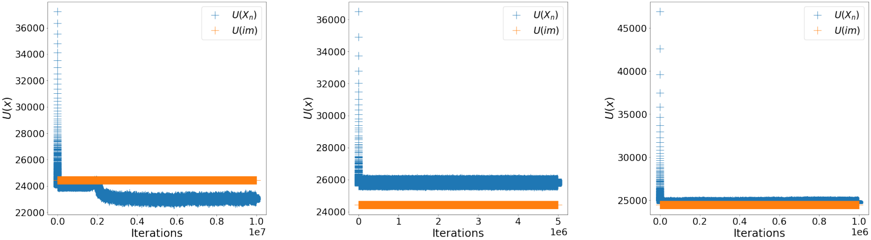
- Prox-GSD posterior potential:

$$\forall x \in \mathbb{R}^d, \; U(x) = \frac{\|Ax - y\|^2}{2\sigma^2} + \frac{1}{2}\|x - N(x, \varepsilon)\|_2^2. \tag{6}$$

Evolution of the posterior potential U_{red} for `Simpson` and `Goldhill` with the SN-DnCNN and FINE induced priors.



Evolution of the exact posterior potential U for `Color Simpson` and `Fox` with PnP-ULA or PPnP-ULA with the Prox-GSD induced prior.

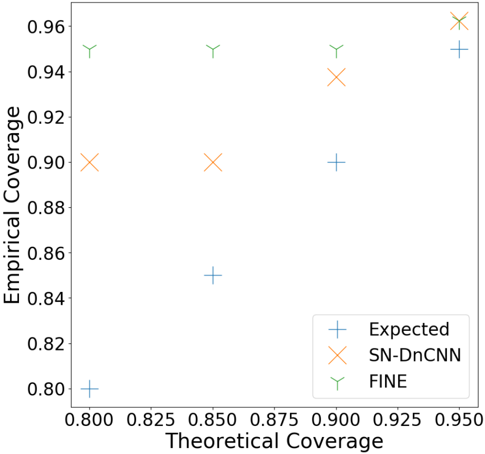


`Color Simpson` PnP-ULA

`Color Simpson` PPnP-ULA

`Fox` PPnP-ULA

Frequentist accuracy of the Bayesian models defined by SN-DnCNN and FINE.



Conclusion.

- Summary.
  - Development of a Langevin based algorithm with detailed convergence guarantees under realistic hypothesis.
  - Analysis of the influence of the plugged prior from a visual and a potential viewpoint.
  - Controlling the Lipschitz constant of the denoiser does not seem to be sufficient for PnP priors.
- Future work.
  - Further investigation of the current PnP priors.
  - Comparison with other methods such as NFs (Hagemann et al., 2022), DDPM based models (Kawar et al., 2022) and Score-matching based methods (Kawar et al., 2021).

Thank you !

If you want to know more you can:

1. ask questions.
2. find the article related to PnP-ULA on Arxiv `https://arxiv.org/abs/2103.04715` and published in *SIAM Imaging Science*.

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. 7).

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. 6).

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. 7).

Hagemann, Paul, Johannes Hertrich, and Gabriele Steidl (2022). *Stochastic Normalizing Flows for Inverse Problems: a Markov Chains Viewpoint*. arXiv: 2109.11375 [cs.LG] (cit. on p. 26).

Hurault, Samuel, Arthur Leclaire, and Nicolas Papadakis (2022). "Gradient Step Denoiser for convergent Plug-and-Play". In: *International Conference on Learning Representations* (cit. on p. 10).

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. 7).

Kawar, Bahjat, Michael Elad, Stefano Ermon, and Jiaming Song (2022). "Denoising diffusion restoration models". In: *arXiv preprint arXiv:2201.11793* (cit. on p. 26).

Kawar, Bahjat, Gregory Vaksman, and Michael Elad (2021). "SNIPS: Solving Noisy Inverse Problems Stochastically". In: *arXiv preprint arXiv:2105.14951* (cit. on p. 26).

Pesquet, Jean-Christophe, Audrey Repetti, Matthieu Terris, and Yves Wiaux (2020). *Learning Maximally Monotone Operators for Image Recovery*. arXiv: 2012.13247 [math.OC] (cit. on p. 10).

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.

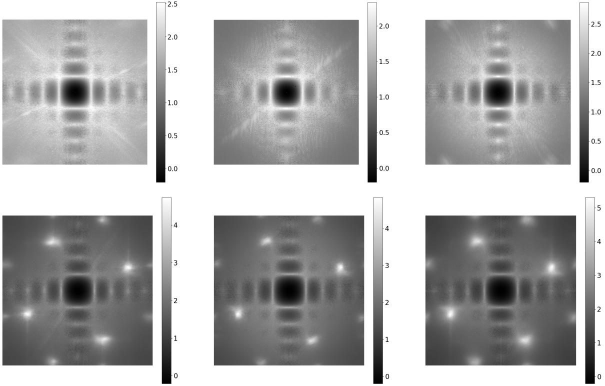
Romano, Yaniv, Michael Elad, and Peyman Milanfar (2017). "The little engine that could: Regularization by denoising (RED)". In: *SIAM Journal on Imaging Sciences* 10.4, pp. 1804–1844 (cit. on pp. 7, 22).

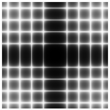
Ryu, Ernest K., Jialin Liu, Sicheng Wang, Xiaohan Chen, Zhangyang Wang, and Wotao Yin (2019). "Plug-and-Play Methods Provably Converge with Properly Trained Denoisers". In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, pp. 5546–5557. arXiv: 1905.05406 (cit. on p. 10).

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

- Hypotheses:
  - $Id - D_\varepsilon$ is Lipschitz and there exists $M : \mathbb{R}^+ \to \mathbb{R}^+$, such that for all $\|x\| \leq R$, $\|D_\varepsilon(x) - D_\varepsilon^*(x)\| \leq M(R)$.
  - The likelihood $p(y|x)$ is bounded, C^1 and $\nabla \log p(y|x)$ is Lipschitz.
  - The MSE loss for D_ε^* under $g_\varepsilon(.|\tilde{x})$ is finite and uniformly bounded.
- Convergence at exponential rate toward $\pi_{\delta,\varepsilon}^C$, non-asymptotic error:

$$|\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_k] - \int_{\mathbb{R}^d} \tilde{x} p(\tilde{x}|y) \mathrm{d}\tilde{x}| \leq C_0\{C_1 \varepsilon^{\beta/4} + C_2 R_C^{-1} + C_3(\sqrt{\delta} + \frac{1}{n\delta} + C_R)\}. \quad (7)$$





Inverse Fourier transform of the blur kernel.

Figure: Log-standard deviation maps in the Fourier domain for the Markov chains defined by PnP-ULA for the deblurring problem. First line: images `Cameraman`, `Simpson`, `Traffic`. Second line: images `Alley`, `Bridge` and `Goldhill`.

Samples of the SN-DnCNN induced prior distribution.