Maximum marginal likelihood estimation of regularisation parameters in Plug & Play Bayesian estimation. Application to non-blind and semi-blind image deconvolution.

Charlesquin Kemajou Mbakam

Joint work with Marcelo Pereyra and Jean-Francois Giovannelli

Maxwell Institue for Mathematical Sciences, Heriot-Watt University

Workshop on Recent Advances in Iterative Reconstruction, London, May 2023

Outline

Imaging inverse problems

- Forward model:

$$y = Hu + w, \tag{1}$$

where,

  - $u \in \mathbb{R}^d$ unknown image, $y \in \mathbb{R}^d$ observed data and $d \in \mathbb{N}$,
  - H a circulant block matrix of dimension $d \times d$ and ...
  - $w \sim \mathcal{N}(0, \sigma^2 Id)$ noise, $\sigma^2 > 0$.

- Deconvolution problems: Estimating u from y.

Recovering u from y is an ill-posed problem $\Longrightarrow \cdots$ need to regularise the solution space.

Play & Play imaging methods

From the Bayes Theorem,

$$\underbrace{p(u|y;\sigma^2)}_{Posterior} \propto \underbrace{p(y|u;\sigma^2)}_{Likelihood}\underbrace{p^*(u)}_{Prior}, \tag{2}$$

Play & Play imaging methods

From the Bayes Theorem,

$$\underbrace{p(u|y;\sigma^2)}_{Posterior} \propto \underbrace{p(y|u;\sigma^2)}_{Likelihood} \underbrace{p^*(u)}_{Prior}, \tag{2}$$

PnP-ULA [Laumont et al., 2021] method proposes a sampling Lagenvin algorithm to draw samples from the Bayesian model

$$\underbrace{p_\epsilon(u|y;\sigma^2)}_{Posterior} \propto \underbrace{p(y|u;\sigma^2)}_{Likelihood} \underbrace{p_\epsilon(u)}_{Prior}, \tag{3}$$

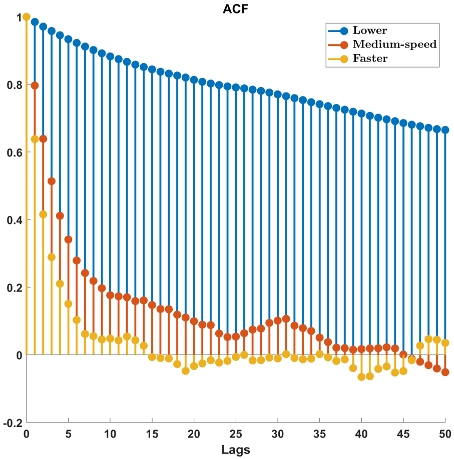
$p_\epsilon(u)$ is the smoothed version of $p^*(u)$ related to a denoiser through Tweedie's identity

$$\epsilon^2 \nabla \log p_\epsilon(u) = ({D_\epsilon}^* u - u) \approx (D_\epsilon u - u)\,. \tag{4}$$

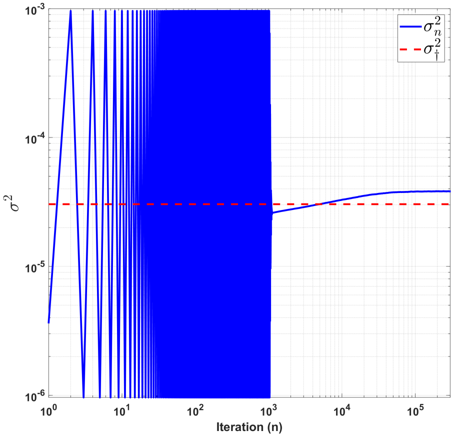
- D_ϵ is an approximation of the optimal MMSE denoiser D_ϵ^*, which recovers u from $\tilde{u} = u + \omega'$ where $\omega' \sim \mathcal{N}(0, \epsilon^2 Id)$.

- PnP-ULA method: ϵ^2 set manually

(a) PnP-ULA: 29.1dB



(b) Autocorrelation



(c) Noise var. σ^2

$\Longrightarrow$ ... the Markov chain generated has poor mixing properties.

$\Longrightarrow$ ... the noise variance of the model is over-estimated.

Main objective

Estimate the regularisation parameter ϵ^2 from the measurement y by computing the MMLE estimation

$$\hat{\epsilon}^2 = \underset{\epsilon^2 \in [\epsilon_0^2, +\infty[}{\operatorname{argmax}} \; p(y|\epsilon^2, \sigma^2), \tag{5}$$

where ϵ_0^2 is a minimum value set a priori.

Note that the marginal likelihood is given by

$$p(y|\epsilon^2, \sigma^2) = \int_{\mathbb{R}^d} p(u, y|\epsilon^2, \sigma^2) du,$$

where

$$p(u, y|\epsilon^2, \sigma^2) = p(y|u, \sigma^2) p_\epsilon(u).$$

Bayesian PnP method in the latent space

- Auxiliary variable

$$u = z + \omega' \qquad \text{where } \omega' \sim \mathcal{N}(0, \rho^2 Id)$$

Then,

$$\epsilon^2 = \epsilon_0^2 + \rho^2, \quad \rho \geq 0,$$

with ϵ_0^2 set a priori.

Estimating ϵ^2 is equivalent to estimating ρ^2

$$\hat{\rho}^2 = \underset{\rho^2 \in [0, +\infty[}{\operatorname{argmax}} \; p(y|\rho^2, \sigma^2).$$

Bayesian PnP method in the latent space

- Auxiliary variable

$$u = z + \omega' \qquad \text{where } \omega' \sim \mathcal{N}(0, \rho^2 Id)$$

Then,

$$\epsilon^2 = \epsilon_0^2 + \rho^2, \quad \rho \geq 0,$$

with ϵ_0^2 set a priori.
Estimating ϵ^2 is equivalent to estimating ρ^2

$$\hat{\rho}^2 = \underset{\rho^2 \in [0, +\infty[}{\operatorname{argmax}} \; p(y|\rho^2, \sigma^2).$$

- Joint probability distribution of u and z

$$p_{\epsilon_0}(u, z|y; \sigma^2, \rho^2) \propto p(y|u; \sigma^2) p(u|z; \rho^2) p_{\epsilon_0}(z) \tag{6}$$

where,

$$p(u|z; \rho^2) \propto \exp\left(||u - z||^2 / (2\rho^2)\right).$$

Benefit of introducing a latent variable

- The likelihood $p(y|z,\rho^2,\sigma^2)$ is strongly log-concave and therefore, running PnP-ULA on z is much faster than running PnP-ULA on u:

$$p(y|z,\rho^2,\sigma^2) = \int_{\mathbb{R}^d} p(y|u;\sigma^2)p(u|z,\rho^2)du$$

- We can easily incorporate additional parameters such as noise variance, and estimate by maximum marginal likelihood estimation

$$\hat{\sigma}^2 = \underset{\sigma^2 \in \Theta_{\sigma^2}}{\operatorname{argmax}}\, p(y|\rho^2,\sigma^2),$$

where Θ_{σ^2} is a convex set of admissible values for σ^2.

Estimation of σ^2 and ρ^2

- We evaluate the Maximum Marginal likelihood estimator from y,

$$(\hat{\sigma}^2, \hat{\rho}^2) \in \underset{\sigma^2 \in \Theta_{\sigma^2}, \rho^2 \in \Theta_{\rho^2}}{\operatorname{argmax}} p(y|\sigma^2, \rho^2).$$

where,

$$p(y|\sigma^2, \rho^2) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} p(y|\tilde{u}, \sigma^2) p(\tilde{u}|\tilde{z}; \rho^2) \pi_{\epsilon_0}(\tilde{z}) d\tilde{u} d\tilde{z}.$$

- In a manner akin to [Vidal et al., 2019], we update ρ^2 and σ^2 given ρ_0^2 and σ_0^2 as follows

$$\rho_{n+1}^2 = \Pi_{\Theta_{\rho^2}} \left[\rho_n^2 + \delta_{n+1} \nabla_{\rho^2} \log p(y|\rho_n^2, \sigma_n^2) \right]$$

and,

$$\sigma_{n+1}^2 = \Pi_{\Theta_{\sigma^2}} \left[\sigma_n^2 + \delta_{n+1} \nabla_{\sigma^2} \log p(y|\rho_n^2, \sigma_n^2) \right]$$

Estimation of σ^2 and ρ^2: Gradients approximation

$$\nabla_{\rho^2} \log p(y|\rho^2, \sigma^2) = -\mathbb{E}_{u,z|y,\rho^2,\sigma^2}\left[\frac{\log p(u|z,\rho^2)}{\rho^2}\right] - \frac{d}{2\rho^2}$$ (Fisher's identity)

Estimation of σ^2 and ρ^2: Gradients approximation

$$\nabla_{\rho^2} \log p(y|\rho^2, \sigma^2) = -\mathbb{E}_{u,z|y,\rho^2,\sigma^2} \left[\frac{\log p(u|z, \rho^2)}{\rho^2} \right] - \frac{d}{2\rho^2} \quad \text{(Fisher's identity)}$$

$$\nabla_{\rho^2} \log p(y|\rho^2, \sigma^2) = -\frac{1}{m} \sum_{k=1}^{m} \left[\frac{\log p(U_k|Z_k, \rho^2)}{\rho^2} \right] - \frac{d}{2\rho^2} \quad \text{(Appro. MC)}$$

$(U_k)_{k=1}^m$ and $(Z_k)_{k=1}^m$ are sampled according to $p(u|y, z; \rho^2, \sigma^2)$ and $p(z|y; \rho^2, \sigma^2)$ respectively.

Estimation of σ^2 and ρ^2: Gradients approximation

$$\nabla_{\rho^2} \log p(y|\rho^2, \sigma^2) = -\mathbb{E}_{u,z|y,\rho^2,\sigma^2} \left[\frac{\log p(u|z, \rho^2)}{\rho^2} \right] - \frac{d}{2\rho^2} \quad \text{(Fisher's identity)}$$

$$\nabla_{\rho^2} \log p(y|\rho^2, \sigma^2) = -\frac{1}{m} \sum_{k=1}^{m} \left[\frac{\log p(U_k|Z_k, \rho^2)}{\rho^2} \right] - \frac{d}{2\rho^2} \quad \text{(Appro. MC)}$$

$(U_k)_{k=1}^m$ and $(Z_k)_{k=1}^m$ are sampled according to $p(u|y, z; \rho^2, \sigma^2)$ and $p(z|y; \rho^2, \sigma^2)$ respectively.

Accordingly,

$$\nabla_{\sigma^2} \log p(y|\rho^2, \sigma^2) = -\frac{1}{m} \sum_{k=1}^{m} \left[\frac{\log p(y|U_k, \sigma^2)}{\sigma^2} \right] - \frac{d}{2\sigma^2}$$

$z \sim p(z|y;\rho^2,\sigma^2)$ and $u \sim p(u|z,y;\rho^2,\sigma^2)$

- Generate $(Z_k)_{k\in\mathbb{N}}$ targetting $p_{\epsilon_0}(z|y;\rho^2,\sigma^2) \propto p(y|z;\rho^2,\sigma^2)\pi_{\epsilon_0}(z)$

$$Z_{k+1} = \Pi_{\mathcal{C}}\left[Z_k + \gamma\nabla_z \log p(y|Z_k,\rho^2) + \gamma\tau \underbrace{\nabla_z \log \pi_{\epsilon_0}(Z_k)}_{(D_{\epsilon_0}Z_k - Z_k)/\epsilon_0^2} + \sqrt{2\gamma}\zeta_{k+1}\right]$$

where $\nabla_z \log p(y|Z_k,\rho^2) = (Z_k - U_k)/\rho^2$ and $\tau > 0$.

- sample $(U_k)_{k\in\mathbb{N}}$ according to $p(u|z,y;\rho^2,\sigma^2)$ which is normal distribution $\mathcal{N}(u;\mu(z,\rho^2,\sigma^2),\Sigma(\rho^2,\sigma^2))$ with

$$\Sigma(\rho^2,\sigma^2) = \left(\frac{H^TH}{\sigma^2} + \frac{I}{\rho^2}\right)^{-1}, \quad \mu(z,\rho^2,\sigma^2) = \Sigma(\rho^2,\sigma^2)\left(\frac{H^Ty}{\sigma^2} + \frac{z}{\rho^2}\right).$$

Therefore, we can update u exactly as follows

$$U_k = \mathbb{E}_{u|z,y;\rho^2,\sigma^2}[u] = \mu(Z_{k+1},\rho^2,\sigma^2)$$

Algorithm

- Sample $(Z_k)_{k\in\mathbb{N}}$ according to $p(z|y;\rho^2,\sigma^2)$ using PnP-ULA

$$Z_{k+1} = \Pi_{\mathcal{C}}\left[Z_k + \gamma\nabla_z \log p(y|Z_k,\rho_k^2,\sigma_k^2) + \tau\gamma\nabla_z \log \pi_{\epsilon_0}(Z_k) + \sqrt{2\gamma}\zeta_{k+1}\right].$$

- Map the latent variable Z_{k+1} to the ambient space

$$X_{k+1} = \mu(Z_{k+1},\rho_{k+1}^2,\sigma_{k+1}^2),$$

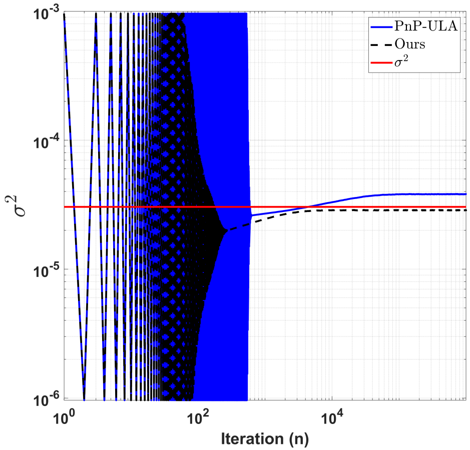
- The parameters ρ_k^2 and σ_k^2 are estimated as follows

$$\rho_{k+1}^2 = \Pi_{\Theta_{\rho^2}}\left[\rho_k^2 - \delta_{k+1}\nabla_{\rho^2}\log p(y|\rho_k^2,\sigma_k^2)\right],$$

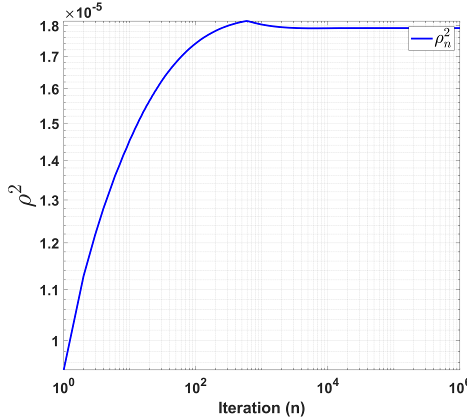
$$\sigma_{k+1}^2 = \Pi_{\Theta_{\sigma^2}}\left[\sigma_k^2 - \delta_{k+1}\nabla_{\sigma^2}\log p(y|\rho_k^2,\sigma_k^2)\right],$$

where $(\delta_t)_{t\in\mathbb{N}}$ is a non-increasing sequence of step-size.

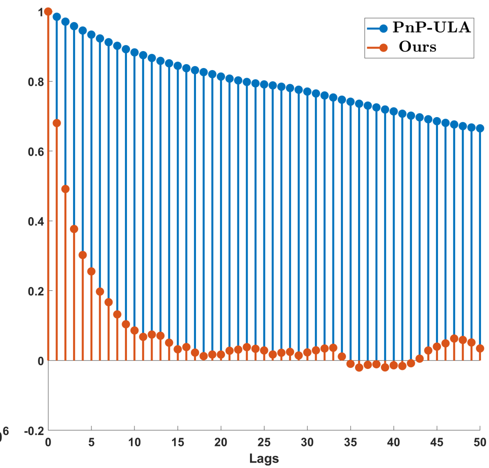
Experiments: Non-blind deblurring



(a) Noise var. σ^2



(b) ρ^2



(c) ACF

$$\hat{\rho}^2_{mmle} = 1.79 \times 10^{-5} > 0$$

Notice that the maximum marginal likelihood estimation of $\rho^2 > 0$ indicates that the original PnP model ($\rho^2 = 0$) is suboptimal.

Experiments: Non-blind deblurring

(a) True u

(b) y: PSNR = 23.5dB

Experiments: Non-blind deblurring

(a) PnP-ULA: PSNR = 29.1dB

(b) Ours: PSNR = 30.0dB

Methods	PSNR	MSE	ESS	Speed-up
PnP-ULA (**PnP-ULA**)	26.16 ± 11.48	$3.3 \times 10^{-3} \pm 6.3 \times 10^{-6}$	3	-
R-PnP-ULA (**Ours**)	$\underline{\mathbf{27.56}} \pm \mathbf{09.02}$	$\underline{\mathbf{2.2}} \times \mathbf{10^{-3}} \pm \mathbf{2.6} \times \mathbf{10^{-6}}$	73	21.37

Conclusion

To conclude:

1. Estimating σ^2 with PnP-ULA leads to an incorrect estimation because the amount of regularisation of the PnP prior is not chosen appropriately.
2. The latent space model generalises the original model, they coincide when $\rho^2 \to 0$

$$p_{\epsilon_0}(x, z|y; \rho^2, \sigma^2) \longrightarrow p_{\epsilon}(x|y; \sigma^2) \quad \text{and} \qquad \epsilon = \epsilon_0$$

3. Notice that the MMLE of $\rho^2 > 0$ indicates that the original model $(\rho^2 = 0)$ is suboptimal.
4. Estimating ρ^2 automatically improves the convergence speed and the reconstruction image in terms of the PSNR.

Thank you very much !!!

Laumont, R., De Bortoli, V., Almansa, A., Delon, J., Durmus, A., and Pereyra, M. (2021).
Bayesian imaging using plug & play priors: when langevin meets tweedie.
arXiv preprint arXiv:2103.04715.

Vidal, A. F., De Bortoli, V., Pereyra, M., and Durmus, A. (2019).
Maximum likelihood estimation of regularisation parameters in high-dimensional inverse problems: an empirical bayesian approach.
arXiv preprint arXiv:1911.11709.