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

Charlesquin Kemajou M.
Joint work with Marcelo Pereyra and Jean-Francois Giovannelli

Maxwell Institue for Mathematical Sciences, Heriot-Watt University

PnP workshop, Paris, December 2022

Outline

Introduction

- 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 $d \times d$ obtained from a blur kernel h and ...
    - $w \sim \mathcal{N}(0, \sigma^2 Id)$ noise, $\sigma^2 > 0$.
- Deconvolution problems: Estimating u from y.

Deconvolution problems can be broadly classified in 3 groups: Non-blind, blind and semi-blind.

Introduction

- Non-blind problems: H is known and $\mathbf{u}$ is unknown. The problem is ill-posed so it requires regularisation to estimate $\mathbf{u}$.

Introduction

- Non-blind problems: H is known and u is unknown. The problem is ill-posed so it requires regularisation to estimate u.

- Blind problems: The operator H and u are completely unknown. The problem is Ill-posed and regularisation on u and h is required.

Introduction

- Non-blind problems: H is known and u is unknown. The problem is ill-posed so it requires regularisation to estimate u.

- Blind problems: The operator H and u are completely unknown. The problem is Ill-posed and regularisation on u and h is required.

- Semi-blind problems: $H \in \mathcal{K}$ where,

$$\mathcal{K} = \left\{ H(\alpha) : \mathbb{R}^d \longrightarrow \mathbb{R}^d, \alpha \in \Theta_\alpha \right\}.$$

  - *Pros:* Introduces more structure.
  - *Cons:* The problem is non-linear w.r.t α

The problem is ill-posed and regularisation on u is required.

Problem formulation in Bayesian framework

- *Likelihood function*

$$p(y|u,\sigma^2) \propto \exp\left(-f_{\sigma^2}(u;y)\right), \tag{2}$$

where $f_{\sigma^2}(u;y) = \frac{||y-Hu||^2}{2\sigma^2}$ is the data fidelity term.

- Maximum likelihood estimation (MLE)

$$\hat{u} = \underset{u\in\mathbb{R}^d}{\operatorname{argmax}} \log p(y|u,\sigma^2)$$

Problem formulation in Bayesian framework

- *Likelihood function*

$$p(y|u,\sigma^2) \propto \exp\left(-f_{\sigma^2}(u;y)\right), \tag{2}$$

where $f_{\sigma^2}(u;y) = \frac{||y-Hu||^2}{2\sigma^2}$ is the data fidelity term.

- Maximum likelihood estimation (MLE)

$$\hat{u} = \underset{u\in\mathbb{R}^d}{\operatorname{argmax}} \log p(y|u,\sigma^2) = \underset{u\in\mathbb{R}^d}{\operatorname{argmin}} ||Hu-y||_2^2 \tag{3}$$

- *Limitations*: MLE achieves limit image reconstruction,
  - H is poorly conditioned
  - The problem is ill-posed (dim u > dim y)

- *Solution*: regularise the solution space.

Problem formulation in Bayesian framework

- *Prior distribution:* $\quad u \sim p(u|\alpha)$,
  where $\alpha^T g(u) = -\log p(u|\alpha)$ with g the regularisation term and α the regularisation parameter.

Problem formulation in Bayesian framework

- *Prior distribution:* $u \sim p(u|\alpha)$,
  where $\alpha^T g(u) = -\log p(u|\alpha)$ with g the regularisation term and α the regularisation parameter.

- *Posterior distribution*

$$p(u|y, \sigma^2, \alpha) \propto p(y|u, \sigma^2) p(u|\alpha), \tag{4}$$

Problem formulation in Bayesian framework

- *Prior distribution:* $\quad u \sim p(u|\alpha)$,
  where $\alpha^T g(u) = -\log p(u|\alpha)$ with g the regularisation term and α the regularisation parameter.

- *Posterior distribution*

$$p(u|y, \sigma^2, \alpha) \propto p(y|u, \sigma^2)p(u|\alpha), \tag{4}$$

- *Retrieve u from y*
  1. *Maximum a posteriori (MAP) estimator*

$$\hat{u}_{MAP} = \underset{u \in \mathbb{R}^d}{\operatorname{argmax}} \log p(u|y, \sigma^2, \alpha) = \underset{u \in \mathbb{R}^d}{\operatorname{argmin}} ||y - Hu||_2^2 + \alpha^T g(u)$$

  2. *Minimum mean squared error (MMSE) estimator*

$$\hat{u}_{MMSE} = \int_{\mathbb{R}^d} \tilde{u} p(\tilde{u}|y, \sigma^2, \alpha) d\tilde{u}$$

How to choose the best prior ?

How to choose the best prior ?

- *Hand-crafted priors model*: Tikhonov regularisation, total variation, ℓ_1 regularisation…
    - Capture the statistical nature of the images
    - We can control α automatically…

Drawback: Doesn't capture probability distribution of the image

How to choose the best prior ?

- *Hand-crafted priors model*: Tikhonov regularisation, total variation, ℓ_1 regularisation...
  - Capture the statistical nature of the images
  - We can control α automatically...

Drawback: Doesn't capture probability distribution of the image

- *Plug and Play priors*
  - Non-local mean, Non-local Bayes and BM3D
  - Machine learning models (capture probability distribution of the image)

Drawback: Requires a training for each level of noise, calibrating α is difficult.

Plug and Play prior

- PnP-ULA [Laumont et al., 2022]

$$X_{k+1} = \Pi_{\mathcal{C}}\left[X_k - \gamma \nabla_x f_{\sigma^2}(X_k; y) + \frac{\alpha\gamma}{\epsilon}\left(D_\epsilon(X_k) - X_k\right) + \sqrt{2\gamma} Z_{k+1}\right]$$

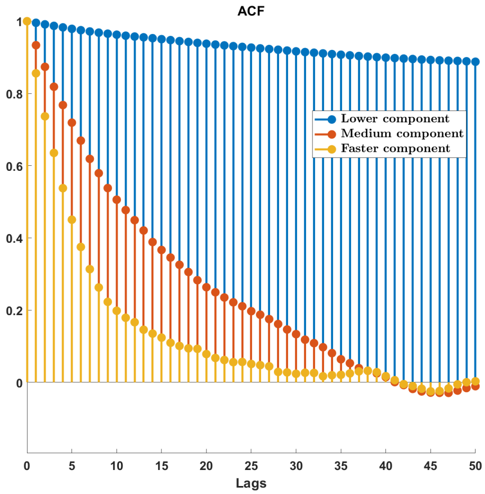
Where,

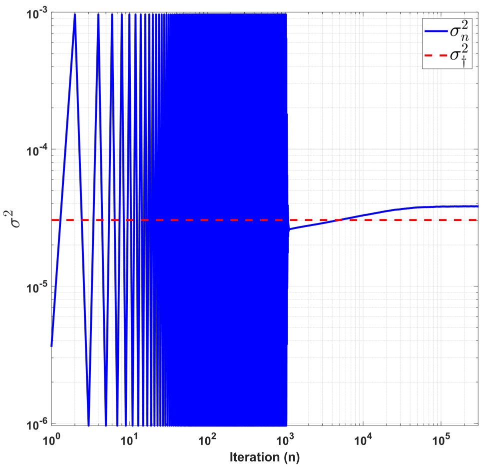
  * α is the regularisation parameter,
  * D_ϵ is the denoiser,
  * ϵ is the noise level.
  * $\Pi_{\mathcal{C}}$ is the projection operator onto the convex set $\mathcal{C}$.

- Limitations of the PnP-ULA:
  * Estimating the variance σ^2 of the model is difficult,
  * The Markov kernel doesn't have good mixing properties.

(a) Blurred

(b) MMSE: 29.1





Regularise PnP-ULA methodology

- Latent variable

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

- Joint probability distribution

$$p(u, z|y; \sigma^2, \rho^2) = \frac{p(y|u; \sigma^2) p(u|z; \rho^2) \pi_{\epsilon_0}(z)}{\int_{\mathbb{R}^{2d}} p(\tilde{u}, \tilde{z}|y; \sigma^2, \rho^2) d\tilde{u} d\tilde{z}} \tag{5}$$

where,

$$p(u|z; \rho^2) \propto \exp\left(\frac{1}{2\rho^2} ||u - z||^2\right)$$

and π_{ϵ_0} is the data distribution with variance ϵ_0^2.

Regularise PnP-ULA methodology

Then simulating u and z from $p(u,z|y;\sigma^2,\rho^2)$ is similar to

1. $u \sim p(u|z,y;\rho^2,\sigma^2) \propto \exp\Big(-\frac{1}{2\sigma^2}||y-Hu||^2 \underbrace{-\frac{1}{2\rho^2}||u-z||^2}_{-\log p(u|z;\rho^2)}\Big)\ ,$

2. $z \sim p(z|u;\rho^2) \propto \exp\big(-\frac{1}{2\rho^2}||u-z||^2\big)\pi_{\epsilon_0}(z).$

It is important to point out that

$$p(u,z|y;\rho^2,\sigma^2) \longrightarrow p(u|y;\sigma^2) \quad \text{when} \qquad \rho \longrightarrow 0$$

The question of interest here, is how to effectively calibrate σ^2 and ρ^2 for the measurement y??

Computation of σ^2 and ρ^2

- We evaluate the Maximum Marginal likelihood estimator from y,

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

where,

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

- Stochastic approximation proximal gradient (SAPG) [Vidal et al., 2019]

$\forall n > 0$,

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

and,

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

Computation of σ^2 and ρ^2

Where,

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

Computation of σ^2 and ρ^2

Where,

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

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

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

Computation of σ^2 and ρ^2

Where,

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

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

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

Accordingly,

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

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

- Simulate u according to $p(u|z,y;\rho^2,\sigma^2)$

$$u \sim p(u|z,y;\rho^2,\sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2}||y-Hu||^2 - \frac{1}{2\rho^2}||u-z||^2\right)$$

$$u \sim \mathcal{N}(u;\mu_u,\Sigma_u)$$

where,

$$\Sigma_u = \left(\frac{H^T H}{\sigma^2} + \frac{I}{\rho^2}\right)^{-1} \text{ and } \mu_u = \Sigma_u\left(\frac{H^T y}{\sigma^2} + \frac{z}{\rho^2}\right)$$

Finally, at iteration $t \geq 0$ we consider

$$u_t = \mathbb{E}_{u|z,y;\rho^2,\sigma^2}[u] = \mu_u$$

$z \sim p(z|u; \rho^2)$

- Simulate z according to $p(z|u; \rho^2)$

$$z \sim p(z|u; \rho^2) \propto \exp\big(-\frac{1}{2\rho^2}||u - z||_2^2\big)\pi_{\epsilon_0}(z).$$

Impossible to sample directly from $p(z|u; \rho^2)$.

- PnP-ULA to simulate z targetting $p(z|y; \rho^2) \propto p(y|z; \rho^2, \sigma^2)\pi_{\epsilon_0}(z)$

$$Z_{t+1} = \Pi_{\mathcal{C}}\left[Z_t - \gamma\nabla_z \log p(y|Z_t, \rho^2) - \gamma\nabla_z \log \pi_{\epsilon_0}(Z_t) + \sqrt{2\gamma}Z_{t+1}\right]$$

where,

$$\nabla_z \log p(y|Z_t, \rho^2) \approx \frac{1}{\rho^2}\mathbb{E}_{u|y,Z_t;\rho^2}[Z_t - u] \qquad \text{(Fisher's identity)}$$

$$\nabla_z \log \pi_{\epsilon_0}(Z_t) \approx \frac{1}{\epsilon_0}(D_{\epsilon_0}Z_t - Z_t) \qquad \text{(Tweedy's identity)}$$

Algorithm 4.1 R-PnP-ULA within SAPG

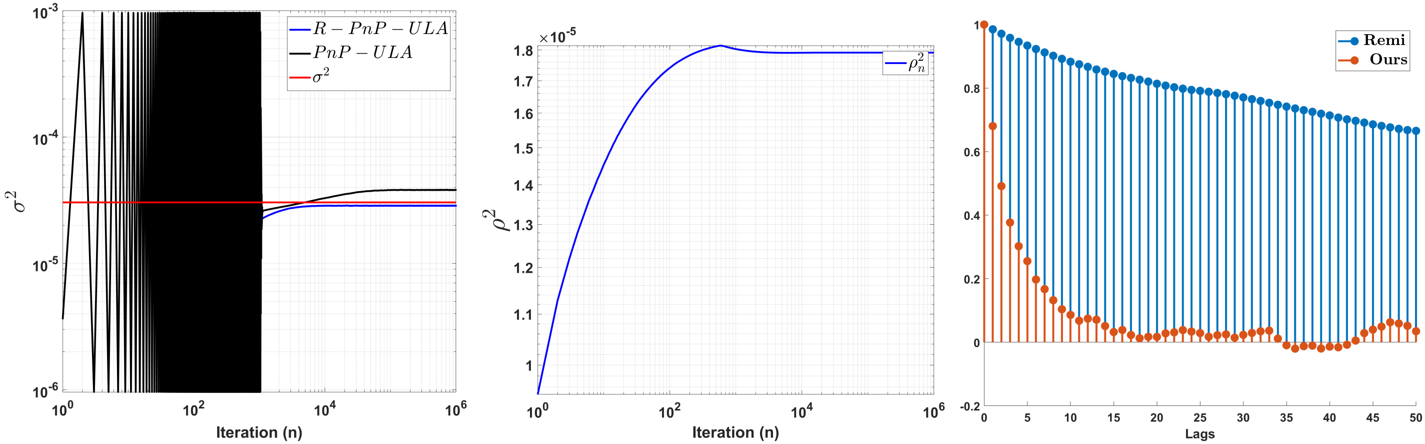
Initialization: $\{\theta_0, U_0^0, Z_0^0\}$, set Θ, define γ, λ and N.
for $n = 0 : N-1$ **do**
  if $n > 0$ **then**
    set $U_0^n = U_{m_n-1}^{n-1}$ and $Z_0^n = Z_{m_n-1}^{n-1}$
  end if
  for $k = 0 : m_n - 1$ **do**
    Sample $U_{k+1}^n \sim p(u|Z_k, y; \theta_n)$
    Sample $\zeta_{k+1} \sim \mathcal{N}(0, Id)$
    $Z_{k+1}^n = Z_k^n + \frac{\gamma}{\rho^2}\left[Z_k^n - \frac{1}{N}\sum_k^N U_k^n\right] + \gamma\alpha\dfrac{D_\epsilon(Z_k^n) - Z_k^n}{\epsilon} + \sqrt{2\gamma}\zeta_{k+1}$
  end for
  $\theta_{n+1} = \Pi_{\Theta_0}\left[\theta_n + \frac{\delta_{n+1}}{m_n}\sum_{k=1}^{m_n}\{\nabla_\theta \log p(U_k^n, Z_k^n|y;\theta)\}\right]$
end for
$\hat{u}_{mmse} = \sum_{n=1}^N \omega_n U_n/\mathcal{N}$ and $\bar{\theta}_N = \sum_{n=1}^N \omega_n\theta_n/\mathcal{N}$, where $\mathcal{N} = \sum_{n=1}^N \omega_n$

Application: Non-blind deblurring...

For a non-blind deconvolution problem,

- Blur kernel: Gaussian of size 9×9 pixels
- Measurement data y of size 512×512 pixels
- Noise level: $30dB$ SNR setup
- Warm-up phase: 5×10^4 iterations
- Sampling phase: 2×10^5 with 30% burn-in.
- Unknown parameters: $\theta = \left(\sigma^2, \rho^2\right)$
- The regularisation parameter of the model: $\alpha = 0.5$
- Denoiser D_{ϵ_0} is Proposed by [Pesquet et al., 2021]
- $\epsilon_0 = \frac{2.25}{255}$
- $\mathcal{C} = [0, 1]^d$.

Illustration. . .



(a) Noise var. σ^2 (b) ρ^2 (c) ACF

Remember,

$$\epsilon_0^2 = (2.25/255)^2 \approx 7.78 \times 10^{-5}$$

and

$$\hat{\rho}^2 = 1.79 \times 10^{-5}$$

We can denote that $\hat{\rho}^2$ is approximately 23% of ϵ_0^2.

(a) Blurred (b) Remi: 29.1 (c) Ours: 29.9

Table 1: Metrics.

Methods	PSNR	MSE	ESS	Speed-up
PnP-ULA (**Remi**)	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 and perspectives...

To conclude,

- Introducing latent variable z permits to:
  - Disentangle the data term from the prior distribution,
  - Regularise the PnP-ULA to achieve state-of-the-art method with good mixing property
  - Efficiently estimate the variance of the model σ^2.
  - Enhance the reconstructed image in terms of PSNR measure.

As perspective,

- Calibrating the regularisation parameter α of the model.

References. . .

Laumont, R., Bortoli, V. D., Almansa, A., Delon, J., Durmus, A., and Pereyra, M. (2022).
Bayesian imaging using Plug & Play priors: when Langevin meets Tweedie.
SIAM Journal on Imaging Sciences, 15(2):701–737.

Pesquet, J.-C., Repetti, A., Terris, M., and Wiaux, Y. (2021).
Learning maximally monotone operators for image recovery.
SIAM Journal on Imaging Sciences, 14(3):1206–1237.

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.