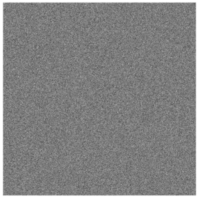
Imaging inverse problems

noise



$u^\star$ $\qquad$ H $\qquad$ y

Imaging inverse problems

- Forward model:

$$y = Hu + w,$$

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 problem: Recovering u from y.

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

Imaging inverse problems

- Non-blind problems: H is known and u is unknown.
- 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
- Blind problems: Both H and u are completely unknown.

These deconvolution problems are ill-posed and additional information about u must be considered.

HERIOT WATT UNIVERSITY

Bayesian Play & Play formulation

From the Bayes Theorem,

$$p(u|y;\sigma^2) \propto p(y|u;\sigma^2)p^*(u)$$

- **Likelihood**: $p(y|u;\sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2}||y - Hu||^2\right)/(2\pi\sigma^2)^{d/2}$
- **Prior**: $p^\star(u) =?$
  1. Handcrafted priors. Total variation prior, sparsity prior, $\cdots$

     $$p^*(u) \propto \exp(-g(u)),$$

     with g being the regularisation term, convex and potentially nonsmooth.

  2. Plug & Play or data driven priors. Image denoising algorithm as a representative of the prior $p^*(u)$.

Play & Play Prior

- $p^*(u)$ has some lower density region.

- Instead, we consider the smooth (noisy) prior

$$p_\epsilon(u) = g_\epsilon * p^\star(u) = \int_{\mathbb{R}^d} g_\epsilon(u-z)p^\star(z)dz,$$

where $g_\epsilon(z)$ is a Gaussian kernel with bandwidth ϵ

$$g_\epsilon(z) = (2\pi\epsilon^2)^{-d/2}\text{exp}\left(-\frac{1}{2\epsilon^2}||z||^2\right).$$

Note that ϵ determines how much we smooth the true prior distribution $p^\star(u)$.

- Depending on the choice of ϵ, $p_\epsilon(u)$ can cover all the lower density region at the expense of some bias.

Bayesian PnP model

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

Why Bayesian Play & Play formulation?

- It relates the prior p_ϵ to a MMSE denoiser $D^\star_\epsilon$

  $$\epsilon^2 \nabla_u \log p_\epsilon(u) = (D^\star_\epsilon u - u) \approx (D_\epsilon u - u)\,. \quad \text{(Tweedie identity)}$$

  Where D_ϵ is an approximation of $D^\star_\epsilon$.
- It combines denoising algorithm and Bayesian inference to achieve high-quality reconstructions.

- PnP approaches applications: image deblurring, image inpainting, super-resolution, compressed sensing and ...

PnP-ULA [Laumont et al., 2022][1] method proposes a sampling Langevin algorithm to draw samples from $p_\epsilon(u|y;\sigma^2)$

$$U_{k+1} = U_k + \gamma \nabla_u \log p(y|U_k, \sigma^2) + \gamma \nabla_u \log p_\epsilon(U_k) + \sqrt{2\gamma}\zeta_{k+1}$$

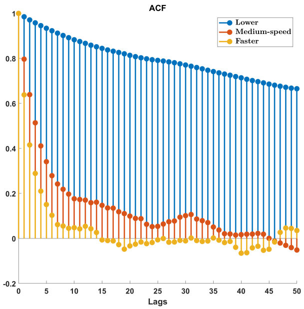
[1] *MS 17 Machine learning models for Bayesian inverse problems* (05.09.23, 2^{nd} *talk*).

Bayesian Play & Play formulation

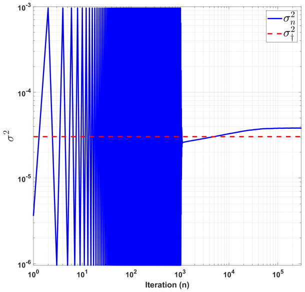
- PnP-ULA method: ϵ^2 set manually

(a) y (23.5dB)

(b) $\hat{u}$ (29.1dB)



(c) ACF



(d) σ^2

- Some Drawbacks of the PnP-ULA.

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

$\Longrightarrow$ The noise variance σ^2 of the model is over-estimated.

Question: Can we improve this !!

HERIOT WATT UNIVERSITY

Bayesian Plug & Play formulation

Main objective

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

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

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

Note that the marginal likelihood is given by

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

Bayesian PnP approach 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 and the noise level ρ^2 unknown.
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

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

where,

$$p(u|z; \rho^2) = g_\epsilon(u - z) = (2\pi\epsilon^2)^{-d/2} \text{exp}\left(-\frac{1}{2\epsilon^2}||u - z||^2\right).$$

Why 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 ρ^2 and σ^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,

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

- Projected gradient descent (PGD) algorithm to update ρ^2 and σ^2

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

and,

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

HERIOT WATT UNIVERSITY

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

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

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

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

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

HERIOT WATT UNIVERSITY

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

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

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

- sample $(U_k)_{k\in\mathbb{N}}$ from $p(u|z,y;\rho^2,\sigma^2)$ which is a normal distribution

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

Therefore, we can update u exactly as follows

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

HERIOT WATT UNIVERSITY

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

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

- The parameters ρ_{k+1}^2 and σ_{k+1}^2 are estimated as follows

$$\rho_{k+1}^2 = \Pi_{\Theta_{\rho^2}}\left[\rho_k^2 - \delta_{k+1}\frac{1}{m}\sum_{i=1}^m\left[\nabla_{\rho^2}\log p(U_i|Z_i,\rho_k^2) - \frac{d}{2\rho_k^2}\right]\right],$$

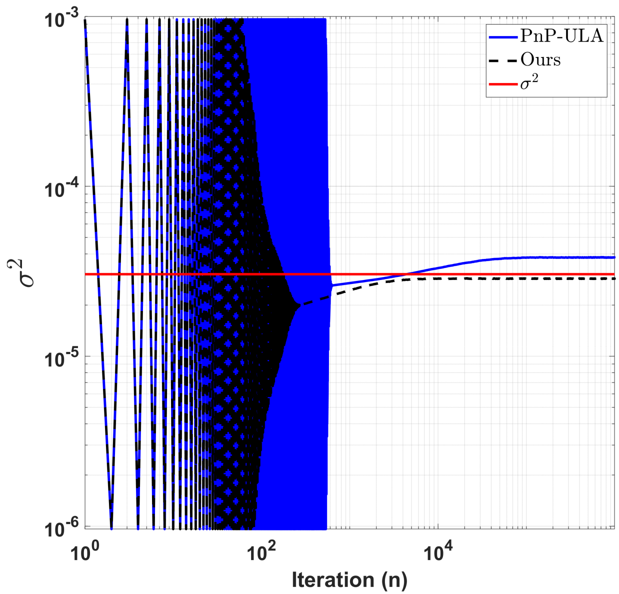
$$\sigma_{k+1}^2 = \Pi_{\Theta_{\sigma^2}}\left[\sigma_k^2 - \delta_{k+1}\frac{1}{m}\sum_{i=1}^m\left[\nabla_{\sigma^2}\log p(y|U_i,\sigma_k^2)\right] - \frac{d}{2\sigma_k^2}\right],$$

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

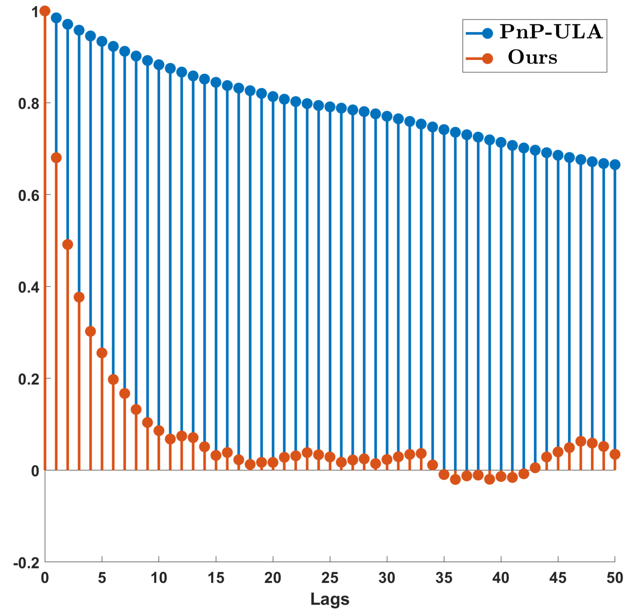
Experiments: Settings

- 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.
- The regularisation parameter of the model: $\alpha = 0.5$
- Denoiser D_{ϵ_0} is Proposed by [Pesquet et al., 2021]
- $\epsilon_0 = \frac{2.25}{255}$
- $\mathcal{C} = [0, 1]^d$.

Experiments: Non-blind deblurring

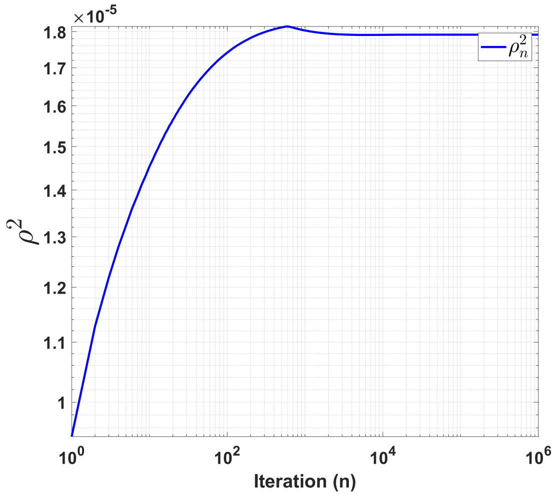


(a) Noise var. σ^2

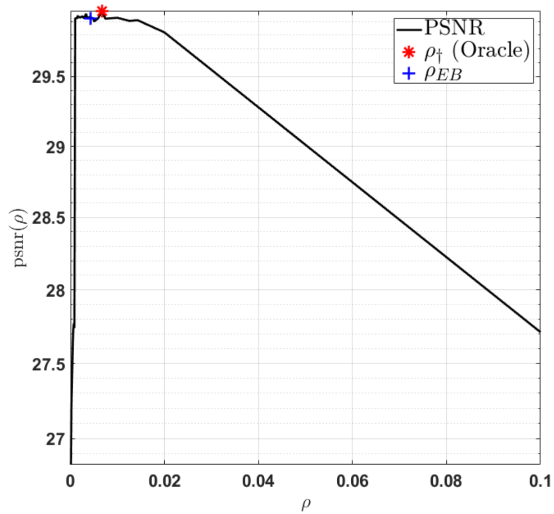


(b) ACF

Experiments: Non-blind deblurring



(a) ρ^2



(b) Calibration (PSNR)

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

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, cont'd

(a) $u^\star$ (b) y (23.5dB)

Experiments: Non-blind deblurring, cont'd

(a) PnP-ULA (29.1dB) (b) Ours (30.0dB)

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

Table 1: Qualitative results over 8 test images.

Experiments: Semi-blind deblurring

Linear model

$$y = H(\alpha, \beta)u + \omega, \quad \omega \sim \mathcal{N}(0, \sigma^2 \mathbb{I})$$

where for any $\alpha > 0, \beta > 0$, $H(\alpha_1, \alpha_2)$ is obtained from the parametric Gaussian blur kernel

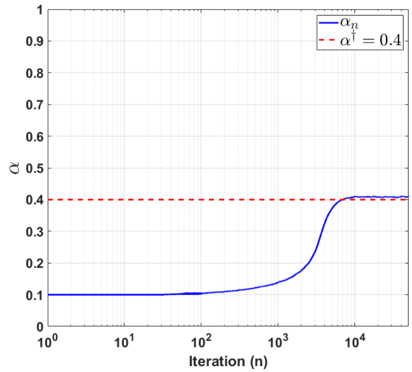
$$h_{\alpha,\beta}(t, v) = \frac{\alpha\beta}{2\pi} \text{exp} \left(-\frac{1}{2} \left(\alpha^2 t^2 + \beta^2 v^2 \right) \right), \quad \forall t, v \in \mathbb{R}.$$

We obtain the semi-blind algorithm by estimating α and β in the non-blind algorithm

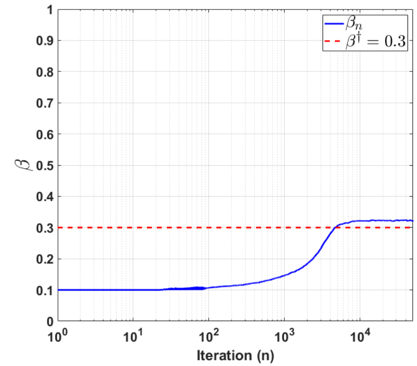
$$\alpha_{k+1} = \Pi_{\Theta_\alpha} \left[\alpha_k - \delta_{k+1} \frac{1}{m} \sum_{i=1}^{m} \nabla_\alpha \log p(y | U_i, \alpha_k, \beta_k, \sigma_k^2) \right]$$

$$\beta_{k+1} = \Pi_{\Theta_\beta} \left[\beta_k - \delta_{k+1} \frac{1}{m} \sum_{i=1}^{m} \nabla_\beta \log p(y | U_i, \alpha_k, \beta_k, \sigma_k^2) \right]$$

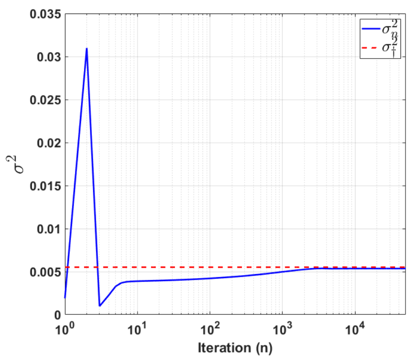
Experiments: Semi-blind deblurring, cont'd



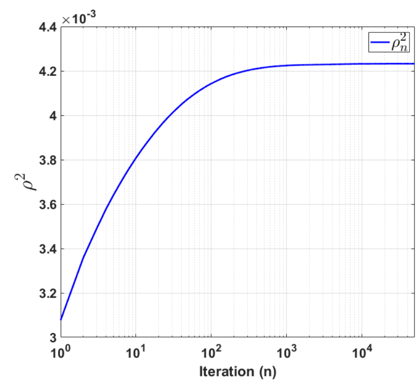
(a) α iterates



(b) β iterates



(c) σ^2 iterates



(d) ρ^2 iterates

HERIOT WATT UNIVERSITY

Experiments: Semi-blind deblurring, cont'd

(a) $u^\star$

(b) y (23.5dB)

(c) $\hat{u}_{\mathsf{MMSE}}$ (29.52dB)

HERIOT WATT UNIVERSITY

Conclusions and perspectives

To conclude:

- Estimating σ^2 with PnP-ULA leads to an incorrect estimation because the amount of regularisation of the PnP prior is not chosen appropriately.
- $p_{\epsilon_0}(x, z|y; \rho^2, \sigma^2) \longrightarrow p_\epsilon(x|y; \sigma^2)$ as $\rho^2 \to 0$
- Notice that the MMLE of $\rho^2 > 0$ indicates that the original model $(\rho^2 = 0)$ is sub-optimal.
- Estimating ρ^2 automatically improves the convergence speed and the reconstruction image in terms of the PSNR.

Next steps

- Study the theoretical convergence of the method.
- Extend this work to PnP diffusion model.

HERIOT WATT UNIVERSITY