Bayesian Inference with Deep Generative Priors Encoded by Neural Networks

Dr. Marcelo Pereyra
http://www.macs.hw.ac.uk/~mp71/

Maxwell Institute for Mathematical Sciences & Heriot-Watt University

UCL, 22 May 2023.

HERIOT WATT UNIVERSITY

Outline

Imaging inverse problems

- We are interested in an unknown image $x^\star \in \mathbb{R}^d$.
- We measure $y \in \mathrm{Y}$, related to $x^\star$ by some mathematical model.
- For example, in many imaging problems

$$y = Ax^\star + w,$$

  for some operator A that is poorly conditioned or rank deficient, and an unknown perturbation or "noise" w.
- The recovery of $x^\star$ from y is usually not well posed. Additional information is required in order to deliver meaningful solutions.

Mathematical imaging frameworks

- There are three main mathematical and computational frameworks for inference in imaging inverse problems:
  1. Mathematical analysis
  2. Bayesian statistics.
  3. Machine learning.
- These frameworks have complementary strengths and weaknesses.
- Our aim is a unifying framework of theory, methods, and algorithms that inherits the benefits of each approach.

Outline

The Bayesian statistical approach

- Model $x^\star$ as a realisation of a r.v. $\mathbb{x}$ on $\mathbb{R}^d$. Use the distribution of $\mathbb{x}$ to regularise the problem and promote expected properties.

- The observation y is a realisation of a r.v. $(\mathbb{y}|\mathbb{x} = x^\star)$.

- Inferences about $x^\star$ from y are derived from the joint distribution of $(\mathbb{x}, \mathbb{y})$ - specified via the decomposition $p(x, y) = p(y|x)p(x)$.

- This determines the posterior distribution, with density

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

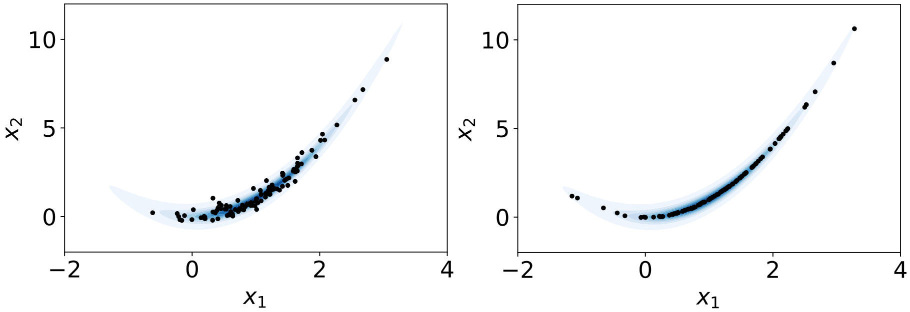
which models our beliefs about $\mathbb{x}$ after observing $\mathbb{y} = y$.

Generative image priors encoded by neural networks

Here, we focus on Bayesian inference based on deep generative priors for problems with abundant training data available to describe $\mathbb{x}$:

1. Let $\{x'_i\}_{i=1}^M$ be a training dataset that represents our prior knowledge about $\mathbb{x}$.
2. We adopt a manifold hypothesis and suppose that $\mathbb{x}$ takes values close to an unknown p-dimensional submanifold of $\mathbb{R}^d$.
3. To estimate the manifold, we introduce a latent r.v. $\mathbb{z}$ on $\mathbb{R}^p$, with $p \ll d$, and a mapping $\nu_\theta : \mathbb{R}^p \mapsto \mathbb{R}^d$, such that the push-forward measure of $\mathbb{z} \sim \mathcal{N}(0, \mathrm{I}_p)$ under ν_θ is close to $\{x'_i\}_{i=1}^M$ (in dist.).
4. We implement ν_θ as a neural network. Can learn ν_θ from $\{x'_i\}_{i=1}^M$ by using, e.g., a VAE, a GAN, or a normalising flow approach.

Illustrative example - Rosenbrock distribution



Left: training data from the two-dimensional Rosenbrock distribution. Right: push-forward of $\mathbb{z} \sim \mathcal{N}(0, \mathrm{I}_p)$ under ν_θ as implemented by a VAE, with $p = 1$.

Posterior distributions for generative priors

- With $\mathbb{z}$ and ν_θ, we have the likelihood $p(y|z) = p(y|x = \nu_\theta(z))$.
- We use Bayes' theorem to derive the posterior for $\mathbb{z}|\mathbb{y} = y$

$$p(z|y) = \frac{p(y|x = \nu_\theta(z))p(z)}{\int_{\mathbb{R}^p} p(y|\tilde{z})p(\tilde{z})\mathrm{d}\tilde{z}} ,$$

- Pushing $(\mathbb{z}|\mathbb{y} = y)$ under $\nu_\theta(z)$ leads to the posterior for $(\mathbb{x}|\mathbb{y} = y)$, which supported on a manifold and does not have a density.

Key questions

Some fundamental questions:

1. Under what conditions on the generative model are the resulting Bayesian models well-posed and amenable to efficient computation? Do the key quantities of interest inherit this well-posed nature?

2. Are these Bayesian methods and algorithms delivering solutions that are meaningful from a non-subjective point of view?

3. Can we perform computation for these models with non-asymptotic accuracy guarantees, under easily verifiable conditions?

In this short talk, we will focus on the first two questions and demonstrate the approach with some numerical experiments.

Key papers

For technical details please see:

1. M. Holden, M. Pereyra, K. Zygalakis, “Bayesian Imaging with Data-Driven Priors Encoded by Neural Networks”, SIAM Journal on Imaging Sciences, 15 (2), 2022. `https://doi.org/10.1137/21M1406313`.

2. S. Melidonis, M. Holden, P. Dobson, Y. Altmann, M. Pereyra, K. Zygalakis, “Empirical Bayesian imaging with conditional generative priors encoded by neural networks”, in preparation.

The oracle Bayesian model

We analyse Bayesian models with data-driven priors in an *M-complete* modelling framework:

- There exists a true - albeit unknown - marginal distribution for $\mathbb{x}$ and posterior distribution for $(\mathbb{x}|\mathbb{y} = y)$.
- Basing inferences on these oracle models is theoretically optimal.
- We henceforth denote this optimal prior distribution by μ. When μ admits a density w.r.t. the Leb. measure on $\mathbb{R}^d$, we denote it by $p^\star$.
- In that case, the posterior for $x|y$ has density

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

Approximating the oracle Bayesian model

- We regard the training data $\{x_i'\}_{i=1}^M$ as a sample from μ.
- When we learn ν_θ and approximate μ by assuming that $\mathbb{x} = \nu_\theta(\mathbb{z})$ for $\mathbb{z} \sim \mathcal{N}(0, \mathrm{I}_p)$, pushing $(\mathbb{z}|\mathbb{y} = y)$ under ν_θ leads to the posterior for $(\mathbb{x}|\mathbb{y} = y)$ that approximates the oracle $p^\star(x|y)$.
- Accurately approximating $p^\star(x|y)$ leads to Bayesian probabilities that map meaningfully to the real-world under a frequentist definition of probability.
- Holden et al. (2022a) establishes that $(\mathbb{z}|\mathbb{y} = y)$ and $(\mathbb{x}|\mathbb{y} = y)$ are well-posed in the sense of Hadamard and have finite moments.

Outline

Illustrative experiments

- We first illustrate the proposed approach with the MNIST dataset.
- We perform the following advanced inferences:
  1. Identify the latent dimension p.
  2. Perform MMSE inference in challenging image denoising, inpainting, and deblurring experiments.
  3. Adopt a likelihood-ratio test to detect out-of-sample observations that should not be analysed with the Bayesian model.
  4. Assess the frequentist accuracy of the Bayesian probabilities reported by the model.
- We report comparisons with MAP estimation under the same model, and with PnP-ADMM by using a DnCNN denoiser.

Identification of manifold dimension p

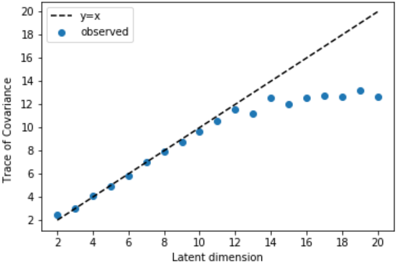
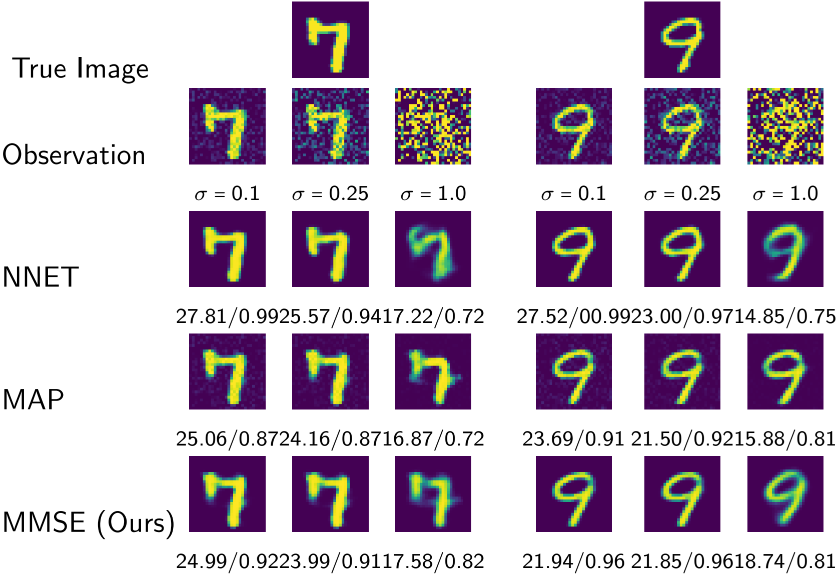


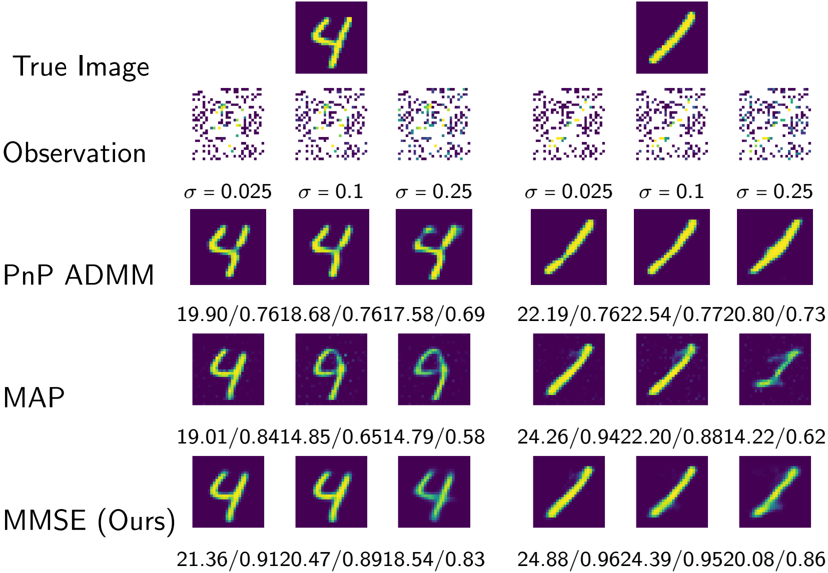
Figure: Trace of sample covariance of $\nu_\theta(x_i)$ across all test images. The amount of information encoded by the prior stabilises for $p \approx 12$, additional dimensions do not significantly increase the amount of prior information encoded .

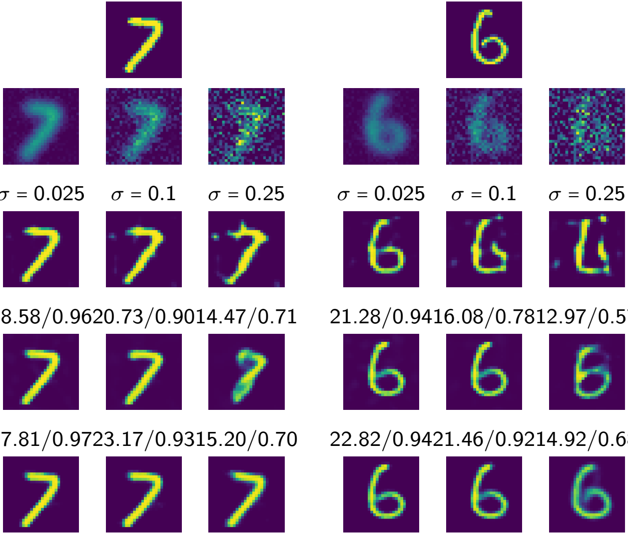
Denoising



True Image
Observation
σ = 0.1 σ = 0.25 σ = 1.0 σ = 0.1 σ = 0.25 σ = 1.0
NNET
27.81/0.99 25.57/0.94 17.22/0.72 27.52/00.99 23.00/0.97 14.85/0.75
MAP
25.06/0.87 24.16/0.87 16.87/0.72 23.69/0.91 21.50/0.92 15.88/0.81
MMSE (Ours)
24.99/0.92 23.99/0.91 17.58/0.82 21.94/0.96 21.85/0.96 18.74/0.81

Inpainting


True Image
Observation
σ = 0.025
σ = 0.1
σ = 0.25
σ = 0.025
σ = 0.1
σ = 0.25
PnP ADMM
19.90/0.76
18.68/0.76
17.58/0.69
22.19/0.76
22.54/0.77
20.80/0.73
MAP
19.01/0.84
14.85/0.65
14.79/0.58
24.26/0.94
22.20/0.88
14.22/0.62
MMSE (Ours)
21.36/0.91
20.47/0.89
18.54/0.83
24.88/0.96
24.39/0.95
20.08/0.86



	$\sigma = 0.025$	$\sigma = 0.1$	$\sigma = 0.25$	$\sigma = 0.025$	$\sigma = 0.1$	$\sigma = 0.25$
True Image						
Observation						
PnP ADMM	28.58/0.96	20.73/0.90	14.47/0.71	21.28/0.94	16.08/0.78	12.97/0.57
MAP	27.81/0.97	23.17/0.93	15.20/0.70	22.82/0.94	21.46/0.92	14.92/0.68
MMSE (Ours)	28.07/0.98	25.08/0.96	24.41/0.95	22.64/0.95	21.21/0.94	17.57/0.78

Likelihood ratio test for out-of-distribution detection

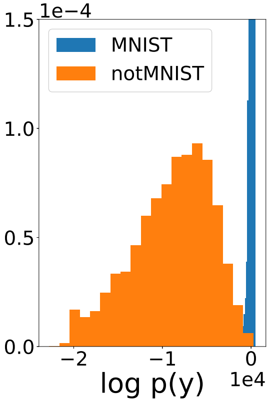


Figure: Denoising

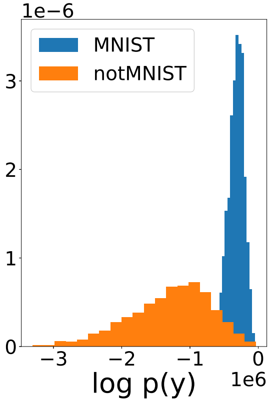


Figure: Inpainting

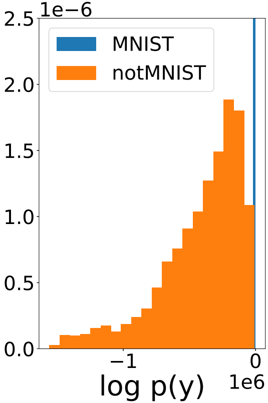


Figure: Deblurring

Figure: Histograms of marginal likelihoods for image denoising, inpainting and deblurring experiments. Out-of-sample detection powers for notMNIST of 99.6%, 88.5% and 99.8% respectively.

Coverage test: frequentist accur. of Bayesian probabilities

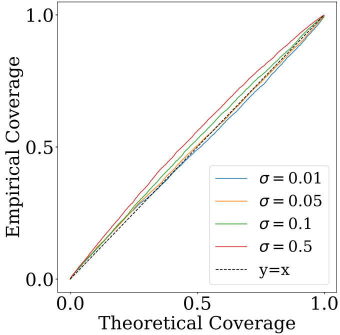


Figure: Denoising

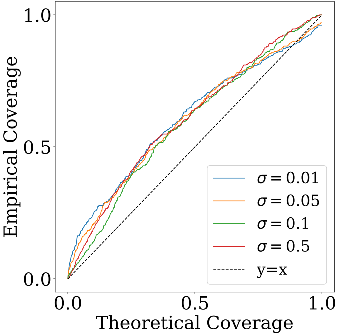


Figure: Inpainting

Outline

Conditional generative priors

- Despite their success in computer vision, scaling generative models to large inference problems reliably is difficult because of mode collapse, spurious modes, or other sources of bias.
- To reduce the difficulty of the machine learning problem, we consider a conditional generative model $\mathbb{x} = \nu_\theta^u(\mathbb{z})$, $\mathbb{z} \sim \mathcal{N}(0, \mathrm{I}_p)$, that models the distribution of $\mathbb{x}$ given some additional r.v. $\mathbb{u}$.
- For this construction to be useful, $\mathbb{u}$ should have low uncertainty given y.

Conditional generative priors

- For example, we let $\mathbb{u}$ denote a low resolution version of $\mathbb{x}$, and implement ν_θ^u by using a normalising flow for image super-resolution.
- This leads to the model

$$p(z|y,u) = \frac{p(y|z,u)p(z)}{p(y|u)}\,,$$

with $p(y|z,u) = p(y|x = \nu_\theta^u(z))$ and $p(y|u) = \int_{\mathbb{R}^p} p(y|\tilde{z},u)p(\tilde{z})\mathrm{d}\tilde{z}$.

Empirical Bayesian imaging with conditional generative priors

- We accurately estimate $u^\star$ from y by maximum marginal likelihood estimation:

$$\hat{u} = \operatorname*{argmax}_{\mu} p_\theta(y|u)\,.$$

- Adopting an empirical Bayesian strategy, we perform inference on $(\mathbb{x}|\mathbb{y} = y, \mathbb{u} = \hat{u})$ by using

$$p(z|y,\hat{u}) = \frac{p(y|z,\hat{u})p(z)}{p(y|\hat{u})}\,,$$

and pushing $(\mathbb{z}|\mathbb{y} = y, \mathbb{u} = \hat{u})$ to $(\mathbb{x}|\mathbb{y} = y, \mathbb{u} = \hat{u})$ by using ν_θ^u.

Bayesian computation

- A simple algorithm to compute $\hat{u}$ probabilities and expectations w.r.t. $p(z|y,\hat{u})$ is the Stochastic Approximation Proximal Gradient scheme

$$Z_{k+1} = Z_k + \delta_k \nabla_z \log p(y|Z_k, u_k) + \delta_k \nabla_z \log p(Z_k) + \sqrt{2\delta_k} Z_{k+1}\,,$$

  and

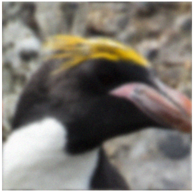
$$u_{k+1} = \Pi_U[u_k + \gamma_k \nabla_u \log p(Z_{k+1}|y, u_k)]\,,$$

  where $Z_{k+1} \sim \mathcal{N}(0, \mathrm{I}_d)$, $(\delta_k)_{k\in\mathbb{N}}$ and $(\gamma_k)_{k\in\mathbb{N}}$ are sequences of step-sizes, and Π_U denotes the Euclidean projection onto the set of admissible values for u.

- This SAPG is reasonably well understood and provably convergent under easily verifiable conditions on $p(z|y)$. See, e.g., `https://doi.org/10.1007/s11222-020-09986-y` and `https://doi.org/10.1137/20M1339829` for details.

Illustrative example - Image deblurring

Recover $x^\star$ from a blurred and noisy measurement y (PSNR, LPIPS).

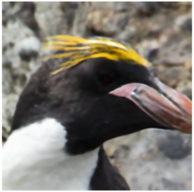


y

$x^\star$

PnP (DnCNN) ADMM
(27.5dB, 0.34)

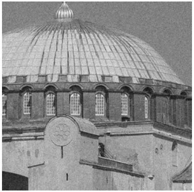
PnP (MMO) FB
(27.8dB, 0.28)



Proposed
(28.2dB, 0.22)

Illustrative example - Image pan-sharpening

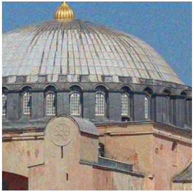
We seek to recover $x^\star$ from two noisy linear observations y_1 and y_2, one with spectral fine details and the other with spatial fine detail.



y_1

y_2

Proposed (31.5dB)



PnP ADMM (28.5dB)

$x^\star$

Outline

Conclusion

- We have studied methodology for Bayesian inference with generative priors encoded by neural networks, learnt from training data.
- Some empirical evidence that the resulting models are sufficiently close to the oracle to report probabilities that are meaningful under a frequentist definition of probability - first example in imaging sciences!
- A key challenge to scale the approach to large problems is that generative models struggle to learn high-dimensional distributions.
- We have addressed that difficulty by adopting an empirical Bayesian approach and considering a conditional generative prior.

Thank you!