Efficient sampling to solve inverse problems with credibility intervals

P. Chainais

N. Dobigeon, F. Cœurdoux (IRIT), M. Vono,
P.-A. Thouvenin, P. Palud, (CRIStAL)
A. Repetti (Heriot-Watt Univ.)

April 5, 2023

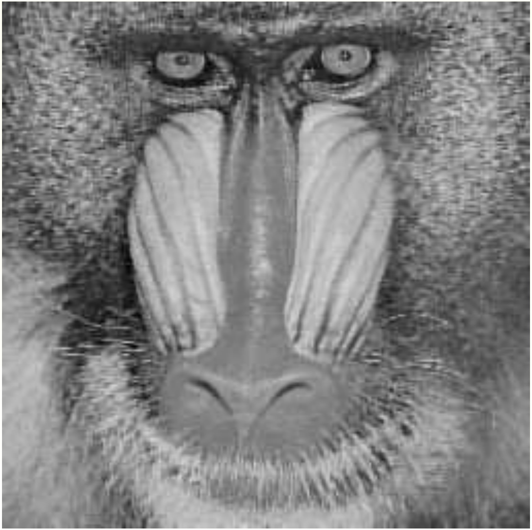
Chaire IA Sherlock

Motivations: inverse problems

Image deblurring

Motivations: inverse problems

Image deblurring



Motivations: inverse problems

Image inpainting

Motivations: inverse problems

Image inpainting

Motivations: inverse problems

Motivations

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

- **solve complex** ill-posed ML or inverse problems
- **big** data in high dimensions
- **good** performances
- **fast** inference algorithms
- **credibility intervals**

with maybe some additional options such as:

- **privacy** preserving
- **distributed** computing

Bayesian approach + MCMC method

(or even better?)

Motivations

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

- **solve complex** ill-posed ML or inverse problems
- **big** data in high dimensions
- **good** performances
- **fast** inference algorithms
- **credibility intervals**

with maybe some additional options such as:

- **privacy** preserving
- **distributed** computing

Bayesian approach + MCMC method

(or even better?)

Flight schedule

Outline

Ill-posed vs well-posed inverse problems

Well-posed problem in the sense of Hadamard

Let $\mathcal{X}$, $\mathcal{Y}$ be two Hilbert spaces. Consider an operator

$$\begin{aligned} \mathcal{A} : \mathcal{X} &\to \mathcal{Y} \\ x &\mapsto \mathcal{A}(x). \end{aligned}$$

Consider the problem which consists in finding x such that $y = \mathcal{A}(x)$. This problem is said to be well-posed in the sense of Hadamard if

1. the problem admits a solution (existence);
2. the problem admits a unique solution (unicity);
3. the solution is stable ($\mathcal{A}^{-1}$ is continuous): for any $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that

$$(\forall y_1, y_2 \in \mathcal{Y}),\ \|y_1 - y_2\| \leq \delta(\varepsilon) \Rightarrow \|x_1 - x_2\| \leq \varepsilon$$

where x_i is a solution to the problem $y_i = \mathcal{A}(x_i)$, $i \in \{1, 2\}$.

Example 1: radio-astronomy - ORION-B

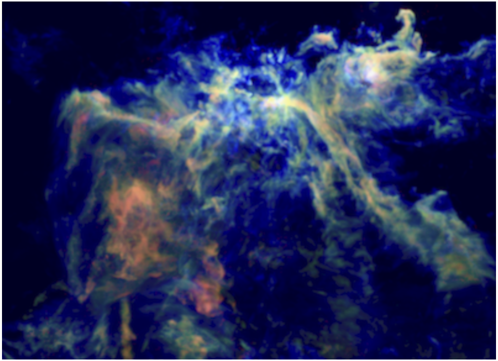
Astrophysics: no ground truth

- observations **y**: radio spectrums w.r.t. chemical composition
- unknowns **x**: physical parameters,

$\Rightarrow$ to understand the birth of stars

Confidence intervals are crucial to acertain predictions

Pierre Palud's PhD with the ORION-B CONSORTIUM



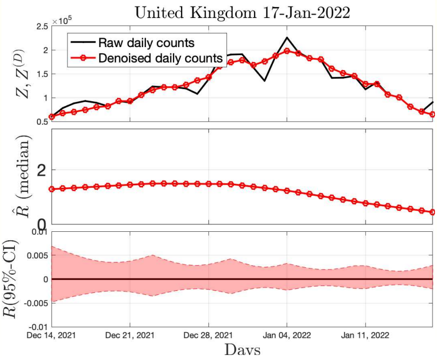
Example 2: estimating the R_0 of Covid-19

Covid-19: no ground truth

- observations **y**: detected contaminations every day
- unknowns **x**: true # of contaminations & R parameter

$\Rightarrow$ to make decisions

Confidence intervals are crucial to acertain predictions



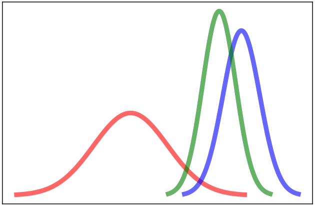
[Abry, Fort, Pascal, Pustelnik 2022]

Bayesian inference[1]

$\mathbf{y}$: available data = observations

$\mathbf{x}$: unknown object of interest

Prior	$\times$	Likelihood	$\longrightarrow$	Posterior
$\mathbf{x} \sim \pi(\mathbf{x})$		$\mathbf{y}\|\mathbf{x} \sim \pi(\mathbf{y}\|\mathbf{x})$		$\mathbf{x}\|\mathbf{y} \sim \pi(\mathbf{x}\|\mathbf{y})$

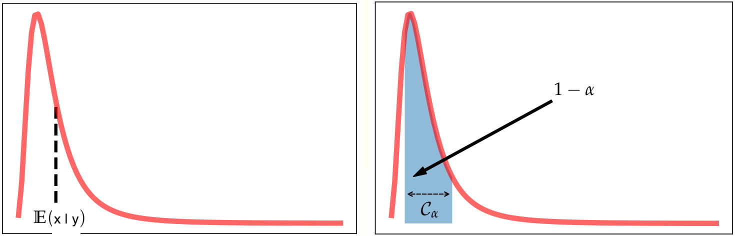


[1] Robert (2001), Gelman et al. (2003)

Bayesian inference

y: available data = observations

x: unknown object of interest



Bayesian estimators

$$\arg\min_{\hat{\mathbf{x}}} \int L(\mathbf{x}, \hat{\mathbf{x}}) \pi(\mathbf{x}|\mathbf{y}) \mathrm{d}\mathbf{x}$$

Credibility regions $\mathcal{C}_\alpha$

$$\int_{\mathcal{C}_\alpha} \pi(\mathbf{x}|\mathbf{y}) \mathrm{d}\mathbf{x} = 1 - \alpha$$

Outline

The usual toolbox of inference

- **Optimization**:
  - problem ⇒ loss function
  - efficient algorithms
  - theoretical guarantees
  - interpretability / functional analysis
- **Bayesian approaches**:
  - probabilitic models
  - uncertainty quantification
- **Machine learning** (deep):
  - adaptive ⇒ relevant
  - outstanding performance

toward the best of all worlds?

The optimization-based approach

Inverse problem $\Rightarrow$ **cost function**

$$\mathbf{y} = \mathcal{A}(\mathbf{x}) + \mathbf{n} \qquad\qquad \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}_{Gauss}$$

$$\Downarrow \qquad\qquad \Downarrow$$

$$f(\mathbf{x}) = f_1(\mathbf{x}|\mathbf{y}) + f_2(\mathbf{x}) \qquad\qquad f(\mathbf{x}) = \frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + f_2(\mathbf{x})$$

$$\Rightarrow \hat{\mathbf{x}} = \arg\min_{\mathbf{x}} f(\mathbf{x})$$

where f is typically

- convex (or not): easy optim., unique solution,
- a sum of various penalties: functional analysis,
- differentiable (or not) $\Rightarrow$ gradient descent (or prox)

The optimization-based approach

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} f_1(\mathbf{x}|\mathbf{y}) + f_2(\mathbf{x})$$

If not differentiable: proximal operators and **splitting**

$$\arg\min_{\mathbf{x}} f_1(\mathbf{x}|\mathbf{y}) + f_2(\mathbf{z}) \text{ such that } \mathbf{x} = \mathbf{z}$$

maybe relaxed to (ADMM)

$$\arg\min_{\mathbf{x},\mathbf{z},\mathbf{u}} f_1(\mathbf{x}|\mathbf{y}) + f_2(\mathbf{z}) + \frac{\alpha}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 + \mathbf{u}^T(\mathbf{x} - \mathbf{z})$$

$$\boxed{\text{prox}_{f_2}(\mathbf{x}) = \arg\min_{\mathbf{z}} f_2(\mathbf{z}) + \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2}$$

$\Longrightarrow$ zoo of prox op.

The Bayesian approach

Inverse problems & **Bayes** posterior $\propto$ likelihood($f1$) $\times$ prior($f2$)

$\Rightarrow$ define a **posterior distribution** $p(\mathbf{x}|\mathbf{y}) \propto p_1(\mathbf{y}|\mathbf{x}) \cdot p_2(\mathbf{x})$

where p_2 is typically

- priors: statistical properties
- conjugate $\Rightarrow$ easy sampling/inference
- log-concave (or not) $\leftrightarrow$ f_2 convex

Solution:

- explicit computations in nice conjugate models
- **sampling methods** and MCMC, e.g. **Gibbs** sampling

$$x_i \sim p(x_i|x_{\setminus i}) \quad \forall 1 \leq i \leq d$$

The Bayesian approach

Conjugate models: the exponential family

Inverse problems & **Bayes** posterior $\propto$ likelihood($f1$) $\times$ prior($f2$)

$\Rightarrow$ define a **posterior distribution** $p(\mathbf{x}|\mathbf{y}) \propto p_1(\mathbf{y}|\mathbf{x}) \cdot p_2(\mathbf{x})$

- **The exponential family** (likelihood)

$$p_1(\mathbf{y}|\mathbf{x}) = h_1(\mathbf{y}) g(\mathbf{x}) \exp\left[\mathbf{x}^T \mathbf{u}(\mathbf{y})\right]$$

- **Conjugate prior** (existence of non-informative priors as well...)

$$p_2(\mathbf{x}|\alpha, \beta) = h_2(\alpha, \beta) g(\mathbf{x})^{\beta} \exp\left[\beta \mathbf{x}^T \alpha\right]$$

- **Posterior distribution** knowing N i.i.d. observations $\mathbf{y}_n$

$$p(\mathbf{x}|\mathbf{Y}) \propto g(\mathbf{x})^{\beta+N} \exp\left[\mathbf{x}^T \left(\sum_n \mathbf{u}(\mathbf{y}_n) + \beta\alpha\right)\right]$$

The Bayesian approach

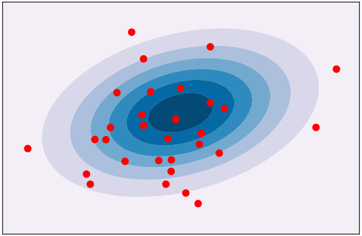
Non conjugate models: sampling and Monte Carlo methods[1]

$$\int h(\mathbf{x})\pi(\mathbf{x}|\mathbf{y})\mathrm{d}\mathbf{x} \approx \frac{1}{N}\sum_{n=1}^{N} h\left(\mathbf{x}^{(n)}\right), \quad \mathbf{x}^{(n)} \sim \pi(\mathbf{x}|\mathbf{y})$$

$$\text{e.g. } \hat{\mathbf{x}}_{MMSE} = \widehat{\mathbf{E}[\mathbf{x}|\mathbf{y}]} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}^{(n)}$$

Sampling challenges: $-\log \pi(\mathbf{x}|\mathbf{y}) = \sum_{i=1}^{b} f_i(\mathbf{x})$

- $\{f_i; i \in [b]\}$: non-conjugate, non-smooth...
- $\mathbf{x} \in \mathbb{R}^d$ with $d \gg 1$



[1] Robert and Casella (2004)

The Bayesian approach: using unchained priors

Non conjugate models: sampling and Monte Carlo methods[2]

Inverse problems & **Bayes** posterior $\propto$ likelihood($f1$) $\times$ prior($f2$)

$\Rightarrow$ define a **posterior distribution** $p(\mathbf{x}|\mathbf{y}) = p_1(\mathbf{x}|\mathbf{y}) \cdot p_2(\mathbf{x})$

If "complex properties"... difficult sampling!

- non-conjugate priors: from optimization, learning,...
- rich models: sophisticated prior distributions
- big datasets: expensive computations
- $f_2 = -\log p_2$ not differentiable

[2] Robert and Casella (2004)

Discretized Langevin process: ULA

Langevin stochastic differential equation:

$$d\mathbf{x}(t) = \nabla \log p(\mathbf{x}(t)|\mathbf{y}) \; + \; \sqrt{2}\, d\mathbf{w}(t),$$

where $\mathbf{w}(t)$ is a d-dimensional Brownian motion.

Unadjusted **L**angevin **A**lgorithm: Euler-Maruyama scheme **(ULA)**

$$\begin{aligned} \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \delta \nabla \log p(\mathbf{x}^{(k)}|\mathbf{y}) + \sqrt{2\delta}\, \mathbf{w}^{(k+1)}, \\ \mathbf{w}^{(k+1)} &\sim \mathcal{N}(0, I_d) \end{aligned}$$

$$\Rightarrow \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta \nabla \underbrace{\log p(\mathbf{y}|\mathbf{x}^{(k)}|\mathbf{y})}_{-f_1(\mathbf{x})} + \delta \nabla \underbrace{\log p(\mathbf{x}^{(k)})}_{-f_2(\mathbf{x})} + \sqrt{2\delta}\, \mathbf{w}^{(k+1)}$$

- discretized Langevin process $\Longrightarrow$ Monte Carlo Markov Chain

Metropolis Adjusted Langevin Algorithm: MALA

Unadjusted Langevin Algorithm: Euler-Maruyama scheme = **ULA**

$$\begin{aligned} \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \delta \nabla \log p(\mathbf{x}^{(k)}|\mathbf{y}) + \sqrt{2\delta}\, \mathbf{w}^{(k+1)}, \\ \mathbf{w}^{(k+1)} &\sim \mathcal{N}(0, I_d) \end{aligned}$$

$\Longrightarrow$ approximation: accuracy vs convergence speed

$\Longrightarrow$ correction by Metropolis-Hastings acceptation step: MALA

Durmus and Moulines (2017)

Rk: SK-ROCK = Runge-Kutta 4 discretization scheme
is much better than Euler-Maruyama
Pereyra et al. (2020)

MYULA: bridging sampling to optimization

- **ULA**: **U**nadjusted **L**angevin **A**lgorithm

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta\nabla \underbrace{\log p(\mathbf{y}|\mathbf{x}^{(k)})}_{-f_1(\mathbf{x})} + \delta\nabla \underbrace{\log p(\mathbf{x}^{(k)})}_{-f_2(\mathbf{x})} + \sqrt{2\delta}\,\mathbf{w}^{(k+1)}$$

$$\mathbf{w}^{(k+1)} \sim \mathcal{N}(0, I_d)$$

$$\Longrightarrow \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \delta\nabla f_1(\mathbf{x}) - \delta\nabla f_2(\mathbf{x}) + \sqrt{2\delta}\,\mathbf{w}^{(k+1)}$$

but $f_2 = -\log p_2$ not differentiable: $\nabla f_2(\mathbf{x}) \Longrightarrow \mathbf{x} - \mathrm{prox}_{\lambda f_2}(\mathbf{x})$

MYULA: bridging sampling to optimization

- **MYULA**: Moreau-Yosida Unadjusted Langevin Algorithm.

  Idea: replace $f_2(\mathbf{x})$ by its Moreau envelope

$$f_2^{(\lambda)}(\mathbf{x}) = \inf_{\mathbf{u}\in\mathbb{R}^d} f_2(\mathbf{u}) + \frac{1}{2\lambda}\|\mathbf{u}-\mathbf{x}\|_2^2$$

$\Longrightarrow \nabla \log p_\lambda$ is Lipshitz continuous: $\nabla {f_2}^{(\lambda)}(\mathbf{x}) = \frac{1}{\lambda}[\mathbf{x} - \mathrm{prox}_{\lambda f_2}(\mathbf{x})]$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} \underbrace{-\,\delta\nabla f_1(\mathbf{x})}_{likelihood} + \delta \underbrace{\frac{1}{\lambda}[\mathrm{prox}_{\lambda f_2}(\mathbf{x}) - \mathbf{x}]}_{prior} + \sqrt{2\delta}\,\mathbf{w}^{(k+1)}$$

Pereyra et al. (2016); Durmus and Moulines (2017); Durmus et al. (2018a) = good approx. when $\lambda \to 0$

Example: radio-astronomy - ORION-B

collab. Obs. of Paris : P. Palud (PhD), F. Le Petit, E. Bron, P.-A. Thouvenin

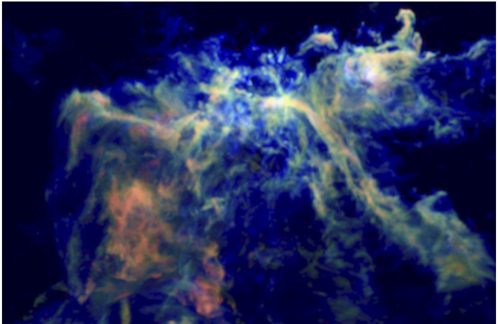
Astrophysics: no ground truth

- observations $\mathbf{y}$: radio spectrums w.r.t. chemical composition
- unknowns $\boldsymbol{\theta}$: physical parameters,

$\Rightarrow$ to understand the birth of stars

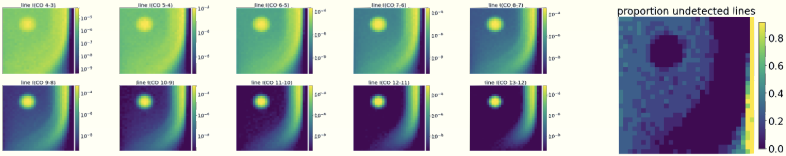
Confidence intervals are crucial to acertain predictions

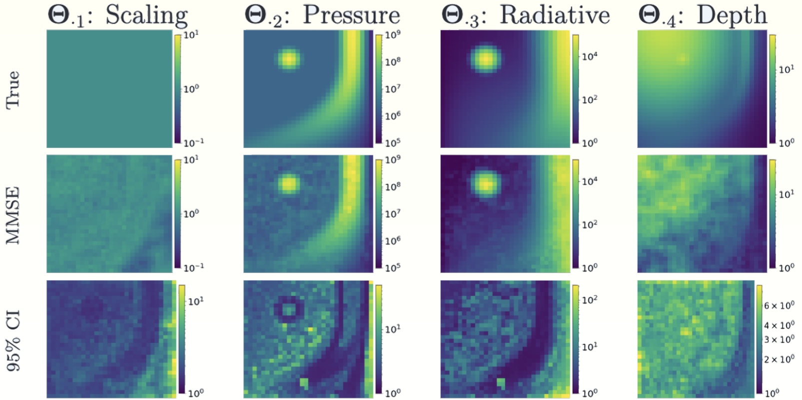
ORION-B CONSORTIUM



Example: radio-astronomy - ORION-B

Synthetic observations $\boldsymbol{Y} \in \mathbb{R}^{900\times 10}$: integrated intensities of excited lines of CO





Example: radio-astronomy - ORION-B

Mixture of noises and sampling non-log-concave posterior distributions

collab. Obs. of Paris : P. Palud (PhD), F. Le Petit, E. Bron, P.-A. Thouvenin

N pixels, L wavelengths, no groundtruth

$$y_{n,\ell} = \max\left\{\omega,\ \epsilon_{n,\ell}^{(m)}\ f_{n,\ell}(\boldsymbol{\Theta}) + \epsilon_{n,\ell}^{(a)}\right\}$$

$\boldsymbol{\theta}_n \in \mathbb{R}^d$	parameters to infer on pixel n
f	black-box, spans multiple decades
$\epsilon_{n,\ell}^{(a)} \sim \mathcal{N}(0, \sigma_a^2)$	instruments noise
$\epsilon_{n,\ell}^{(m)} \sim \log\mathcal{N}(0, \sigma_m^2)$	calibration error
$\omega > 0$	instrument detectability limit

How to deal with | black-box and non linear forward map f ?

mixture of additive and multiplicative noises?

Example: radio-astronomy - ORION-B

Mixture of noises and sampling non-log-concave posterior distributions

collab. Obs. of Paris : P. Palud (PhD), F. Le Petit, E. Bron, P.-A. Thouvenin

N pixels, L wavelengths, no groundtruth

$$y_{n,\ell} = \max \left\{ \omega, \; \epsilon_{n,\ell}^{(m)} \; f_{n,\ell}(\boldsymbol{\Theta}) + \epsilon_{n,\ell}^{(a)} \right\}$$

$\boldsymbol{\theta}_n \in \mathbb{R}^d$	parameters to infer on pixel n
f	black-box, spans multiple decades
$\epsilon_{n,\ell}^{(a)} \sim \mathcal{N}(0, \sigma_a^2)$	instruments noise
$\epsilon_{n,\ell}^{(m)} \sim \log \mathcal{N}(0, \sigma_m^2)$	calibration error
$\omega > 0$	instrument detectability limit

How to deal with | black-box and non linear forward map f ?

mixture of additive and multiplicative noises?

Example: radio-astronomy - ORION-B

Mixture of noises and sampling non-log-concave posterior distributions

collab. Obs. of Paris : P. Palud (PhD), F. Le Petit, E. Bron, P.-A. Thouvenin

N pixels, L wavelengths, no groundtruth

$$y_{n,\ell} = \max \left\{ \omega,\ \epsilon_{n,\ell}^{(m)}\ f_{n,\ell}(\boldsymbol{\Theta}) + \epsilon_{n,\ell}^{(a)} \right\}$$

$\boldsymbol{\theta}_n \in \mathbb{R}^d$	parameters to infer on pixel n
f	black-box, spans multiple decades
$\epsilon_{n,\ell}^{(a)} \sim \mathcal{N}(0, \sigma_a^2)$	instruments noise
$\epsilon_{n,\ell}^{(m)} \sim \log \mathcal{N}(0, \sigma_m^2)$	calibration error
$\omega > 0$	instrument detectability limit

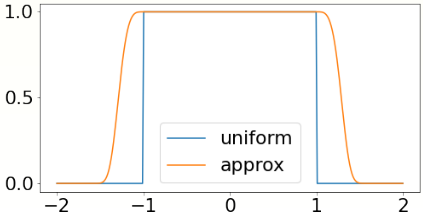
How to deal with
- black-box and non linear forward map f ?
- mixture of additive and multiplicative noises?

A priori & regularization

a priori information on $\boldsymbol{\Theta} \in \mathbb{R}^{N \times D}$ combines 2 priors:

- **spatial regularization**, e.g.,
  - smoothed Total Variation (TV is not diff. $\Rightarrow$ MYULA)
  - L_2-norm of image gradient
  - L_2-norm of image Laplacian
  - L_2-norm of image wavelet decomposition

- **validity domain** for each physical parameter $\theta_{n,d}$
  $\Longrightarrow$ BUT non-smooth
  $\Longrightarrow$ smooth penalty function when $\theta_{n,d}$ is out of validity domain:



Example: radio-astronomy - ORION-B

Proposed sampler: mixing 2 kernels

- Forward model covers multiple decades

  → Preconditioned-MALA kernel with RMSProp

  Role: Efficient local exploration
  Limitation: restricted to smooth log-posteriors

- Non-log-concave posterior

  → Multiple-Try Metropolis (MTM) kernel

  Role: jumps between modes

Illustration: 2D Gaussian mixture model - MALA steps

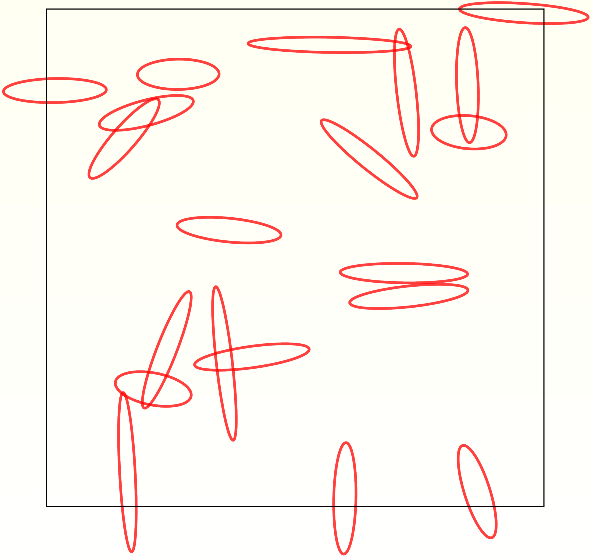


Illustration: 2D Gaussian mixture model - MALA steps

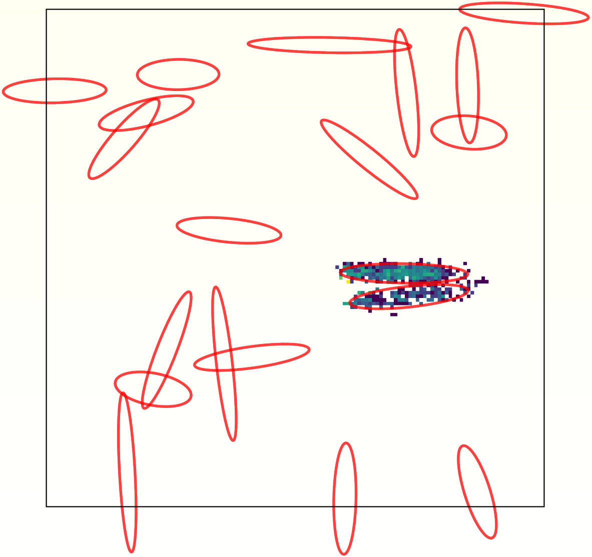


Illustration: 2D Gaussian mixture model - MALA steps

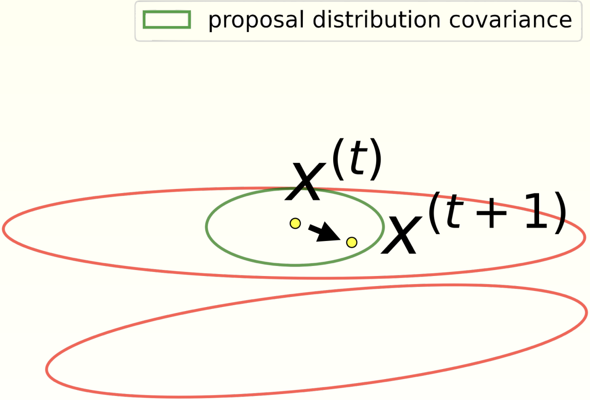


Illustration: 2D Gaussian mixture model - MTM steps

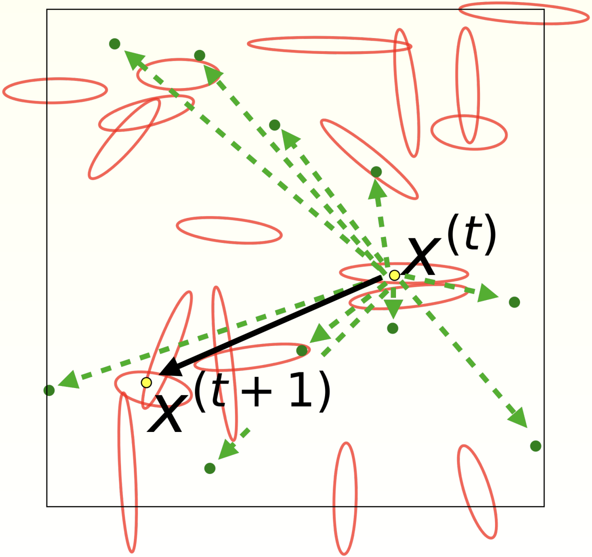


Illustration: 2D Gaussian mixture model - MTM steps

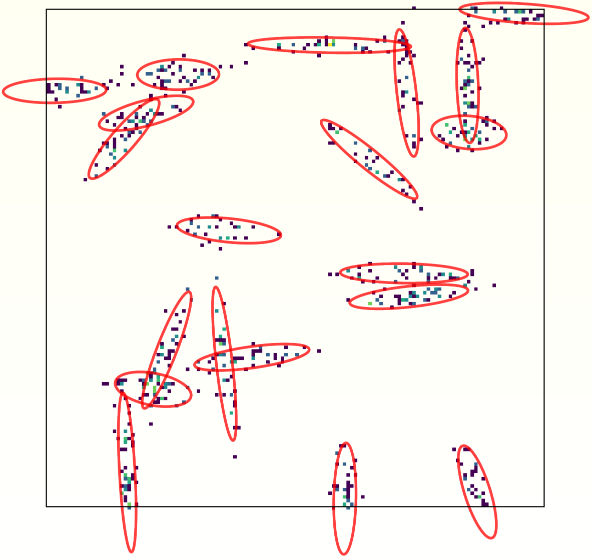
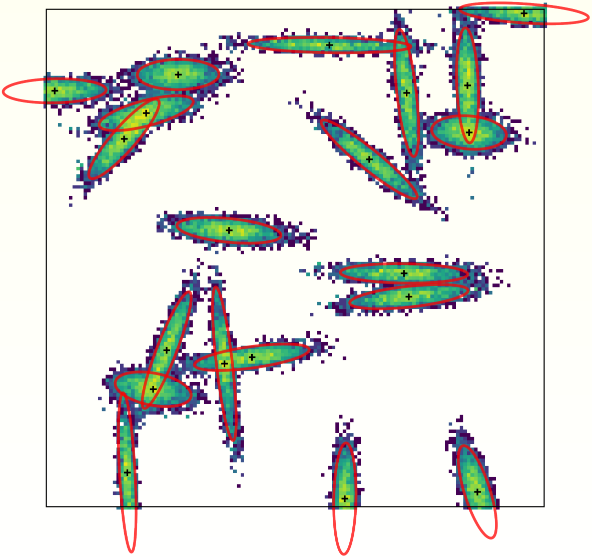
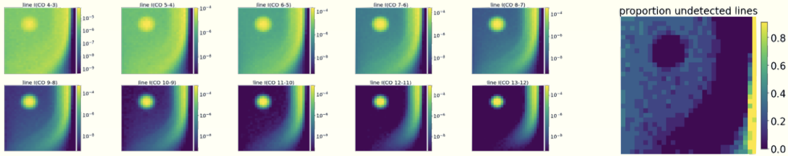


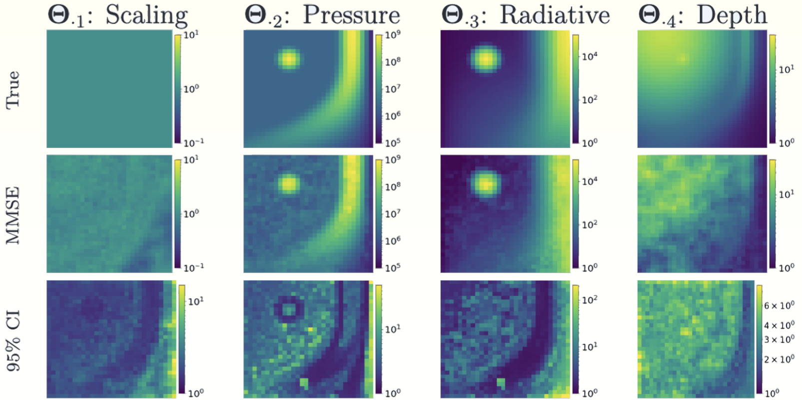
Illustration: 2D Gaussian mixture model - MALA + MTM



Application to a synthetic dataset

Synthetic observations $\boldsymbol{Y} \in \mathbb{R}^{900\times 10}$: integrated intensities of excited lines of CO



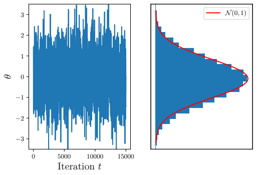


Outline

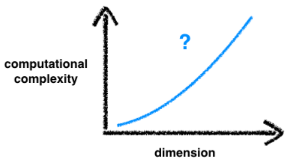
The Bayesian approach augmented by splitting

Exploit the synergy: Monte Carlo sampling / optimization

Seminal works : HMC, (MY)ULA



- efficient & simple sampling
- in high dimension
- in distributed architectures



Recall: $\text{prox}_{\lambda f_2}(x) = \arg\min_{\mathbf{z}} \lambda f_2(\mathbf{z}) + \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 \Rightarrow$ zoo of prox

The Bayesian approach augmented by splitting: AXDA

Inverse problems & **Bayes** posterior $\propto$ likelihood($f1$) $\times$ prior($f2$)

$\Rightarrow$ define a **posterior distribution** $p(\mathbf{x}|\mathbf{y}) = p_1(\mathbf{x}|\mathbf{y}) \cdot p_2(\mathbf{x})$

The Bayesian approach augmented by splitting: AXDA

Inverse problems & **Bayes**   posterior $\propto$ likelihood($f1$) $\times$ prior($f2$)

$\Rightarrow$ define a **posterior distribution** $p(\mathbf{x}|\mathbf{y}) = p_1(\mathbf{x}|\mathbf{y}) \cdot p_2(\mathbf{x})$

The Bayesian approach augmented by splitting: AXDA

Inverse problems & **Bayes** posterior $\propto$ likelihood($f1$) $\times$ prior($f2$)

$\Rightarrow$ define a **posterior distribution** $p(\mathbf{x}|\mathbf{y}) = p_1(\mathbf{x}|\mathbf{y}) \cdot p_2(\mathbf{x})$

"complex properties" $\Longrightarrow$ difficult sampling

Strategy: **Divide-to-Conquer** + **efficient sampling**

$\Longrightarrow$ **splitting (SP)** and **augmentation (SPA)**

Approximate the true posterior: **A**symp. e**X**act **D**ata **A**ugment.

$$\pi(\mathbf{x}) \propto \exp\left[-f_1(\mathbf{x}) - f_2(\mathbf{x})\right]$$
$$\Downarrow$$
$$\pi_\rho(\mathbf{x}, \mathbf{z}, \mathbf{u}) \propto \exp\left[-f_1(\mathbf{x}) - f_2(\mathbf{z}) - \frac{1}{2\rho^2}\|\mathbf{u} - \mathbf{x} + \mathbf{z}\|_2^2 - \frac{1}{2\alpha^2}\|\mathbf{u}\|^2\right]$$

Recall: $\text{prox}_{\lambda f_2}(x) = \arg\min_{\mathbf{z}} \lambda f_2(\mathbf{z}) + \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2$

The Bayesian approach augmented by splitting: AXDA

Inverse problems & **Bayes** posterior $\propto$ likelihood($f1$) $\times$ prior($f2$)

$\Rightarrow$ define a **posterior distribution** $p(\mathbf{x}|\mathbf{y}) = p_1(\mathbf{x}|\mathbf{y}) \cdot p_2(\mathbf{x})$

"complex properties" $\Longrightarrow$ difficult sampling

Strategy: **Divide-to-Conquer** + **efficient sampling**

$\Longrightarrow$ **splitting (SP)** and **augmentation (SPA)**

Approximate the true posterior: **A**symp. e**X**act **D**ata **A**ugment.

$$\pi(\mathbf{x}) \propto \exp\left[-f_1(\mathbf{x}) - f_2(\mathbf{x})\right]$$
$$\Downarrow$$
$$\pi_\rho(\mathbf{x}, \mathbf{z}, \mathbf{u}) \propto \exp\left[-f_1(\mathbf{x}) - f_2(\mathbf{z}) - \frac{1}{2\rho^2}\|\mathbf{u} - \mathbf{x} + \mathbf{z}\|_2^2 - \frac{1}{2\alpha^2}\|\mathbf{u}\|^2\right]$$

Recall: $\mathrm{prox}_{\lambda f_2}(x) = \arg\min_{\mathbf{z}} \lambda f_2(\mathbf{z}) + \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2$

The Bayesian approach augmented by splitting: AXDA

Inverse problems & **Bayes** posterior $\propto$ likelihood($f1$) $\times$ prior($f2$)

$\Rightarrow$ define a **posterior distribution** $p(\mathbf{x}|\mathbf{y}) = p_1(\mathbf{x}|\mathbf{y}) \cdot p_2(\mathbf{x})$

"complex properties" $\Longrightarrow$ **difficult sampling**

Strategy: **Divide-to-Conquer** + **efficient sampling**

$\Longrightarrow$ **splitting (SP)** and **augmentation (SPA)**

Approximate the true posterior: **A**symp. e**X**act **D**ata **A**ugment.

$$\pi(\mathbf{x}) \propto \exp\left[-f_1(\mathbf{x}) - f_2(\mathbf{x})\right]$$
$$\Downarrow$$
$$\pi_\rho(\mathbf{x}, \mathbf{z}, \mathbf{u}) \propto \exp\left[-f_1(\mathbf{x}) - f_2(\mathbf{z}) - \frac{1}{2\rho^2}\|\mathbf{u} - \mathbf{x} + \mathbf{z}\|_2^2 - \frac{1}{2\alpha^2}\|\mathbf{u}\|^2\right]$$

Recall: $\text{prox}_{\lambda f_2}(x) = \arg\min_{\mathbf{z}} \lambda f_2(\mathbf{z}) + \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2$

Let $\pi \in L^1$ a target **probability distribution** with density with respect to (w.r.t.) the Lebesgue measure

$$\boxed{\pi(\mathbf{x}) \propto \exp(-f(\mathbf{x}))}$$

where $f : \mathcal{X} \subseteq \mathbb{R}^d \to (-\infty, +\infty]$ stands for a **potential** function.

With a slight abuse of notations, π shall refer to

- a prior $\pi(\mathbf{x})$,
- a likelihood $\pi(\mathbf{x}) \triangleq \pi(\mathbf{y}|\mathbf{x})$,
- a posterior $\pi(\mathbf{x}) \triangleq \pi(\mathbf{x}|\mathbf{y})$,

where $\mathbf{y}$ are observations.

Asymptotically exact data augmentation (AXDA)
Motivations

Let $\pi \in L^1$ a target **probability distribution** with density with respect to (w.r.t.) the Lebesgue measure

$$\boxed{\pi(\mathbf{x}) \propto \exp(-f(\mathbf{x}))}$$

where $f : \mathcal{X} \subseteq \mathbb{R}^d \to (-\infty, +\infty]$ stands for a **potential** function.

Assumption 1

Inference from π is difficult and possibly inefficient.

Examples:

- non-trivial maximum likelihood estimation
- difficult posterior sampling with poor mixing chains

Data augmentation (DA)

Idea: introduce auxiliary variables $\mathbf{z}$ such that

$$\int_{\mathcal{Z}} \pi(\mathbf{x}, \mathbf{z}) \mathrm{d}\mathbf{z} = \pi(\mathbf{x}).$$

Numerous well-known **advantages**:

- augmented likelihood $\pi(\mathbf{x}, \mathbf{z}) \triangleq \pi(\mathbf{y}, \mathbf{z}|\mathbf{x})$ **easier** to work with
- joint posterior $\pi(\mathbf{x}, \mathbf{z}) \triangleq \pi(\mathbf{x}, \mathbf{z}|\mathbf{y})$ with **simpler** conditionals
- **improved** inference (multimodal problems, mixing properties)

The art of exact data augmentation: XDA

Unfortunately, satisfying

$$\int_{\mathcal{Z}} \pi(\mathbf{x}, \mathbf{z}) \mathrm{d}\mathbf{z} = \pi(\mathbf{x}) \quad \text{(XDA)}$$

is a matter of **art** (van Dyk and Meng 2001).

Difficulties:

- **finding** $\pi(\mathbf{x}, \mathbf{z})$ (Geman and Yang 1995)
- **scaling** in high-dimensional/big data settings (Neal 2003; Polson et al. 2013).

Idea: relax (XDA) while keeping XDA's advantages

How to build $\pi_\rho(\mathbf{x}, \mathbf{z})$ such that $\int \pi_\rho(\mathbf{x}, \mathbf{z}) \mathrm{d}\mathbf{z} \xrightarrow[\rho \to 0]{} \pi(\mathbf{x})$?

Asymptotically exact data augmentation (AXDA)

Let consider an augmented density $p_\rho(\mathbf{x}, \mathbf{z})$ and define

$$\pi_\rho(\mathbf{x}) = \int_{\mathcal{Z}} p_\rho(\mathbf{x}, \mathbf{z}) \mathrm{d}\mathbf{z},$$

where $\rho > 0$.

Assumption 2

For all $\mathbf{x} \in \mathcal{X}$, $\lim_{\rho \to 0} \pi_\rho(\mathbf{x}) = \pi(\mathbf{x})$.

Theorem 1 (Scheffé 1947)

Under Assumption 2,

$$\|\pi_\rho - \pi\|_{\mathrm{TV}} \xrightarrow[\rho \to 0]{} 0.$$

Choice of the augmented density

Take inspiration from variable splitting in optimization (Boyd et al. 2011)...

This motivates the choice (Vono et al. 2019a)

$$p_\rho(\mathbf{x}, \mathbf{z}) \propto \exp(-f(\mathbf{z}) - \phi_\rho(\mathbf{x}, \mathbf{z}))$$

- **simplify** the inference (Vono et al. 2019a)
- **distribute** the inference (Rendell et al. 2021)
- **accelerate** the inference (Vono et al. 2019a).

Splitted Gibbs sampling (SGS)

$$\pi(\mathbf{x}) \propto \exp\left[-f_1(\mathbf{x}) - f_2(\mathbf{x})\right]$$

$$\Downarrow$$

$$\pi(\mathbf{x}, \mathbf{z} | \mathbf{x} = \mathbf{z}) \propto \exp\left[-f_1(\mathbf{x}) - f_2(\mathbf{z})\right] \text{ knowing that } \mathbf{x} = \mathbf{z}$$

$$\Downarrow$$

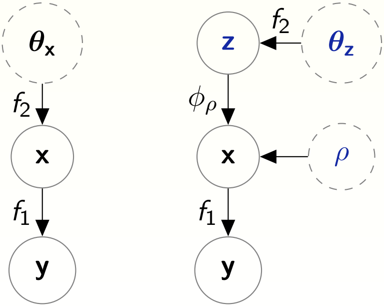
$$\pi_\rho(\mathbf{x}, \mathbf{z}) \propto \exp\left[-f_1(\mathbf{x}) - f_2(\mathbf{z}) - \frac{1}{2\rho^2}\|\mathbf{x} - \mathbf{z}\|_2^2\right]$$

Splitted Gibbs sampling (SGS)

$$\pi(\mathbf{x}) \propto \exp\left[-f_1(\mathbf{x}) - f_2(\mathbf{x})\right]$$

$$\Downarrow$$

$$\pi_\rho(\mathbf{x}, \mathbf{z}) \propto \exp\left[-f_1(\mathbf{x}) - f_2(\mathbf{z}) - \phi_\rho(\mathbf{x}, \mathbf{z})\right]$$



Splitted Gibbs sampling (SP): Theorem

Consider the marginal of $\mathbf{x}$ under π_ρ:

$$p_\rho(\mathbf{x}) = \int_{\mathbb{R}^d} \pi_\rho(\mathbf{x}, \mathbf{z}) \mathrm{d}\mathbf{z} \propto \int_{\mathbb{R}^d} \exp\left[-f_1(\mathbf{x}) - f_2(\mathbf{z}) - \phi_\rho(\mathbf{x}, \mathbf{z})\right] \mathrm{d}\mathbf{z} \ .$$

Theorem

Assume that in the limiting case $\rho \to 0$, ϕ_ρ is such that

$$\frac{\exp\left(-\phi_\rho(\mathbf{x}, \mathbf{z})\right)}{\int_{\mathbb{R}^d} \exp\left(-\phi_\rho(\mathbf{x}, \mathbf{z})\right) \mathrm{d}\mathbf{x}} \xrightarrow[\rho \to 0]{} \delta_{\mathbf{x}}(\mathbf{z})$$

Then p_ρ coincides with π when $\rho \to 0$, that is

$$\|p_\rho - \pi\|_{\mathrm{TV}} \xrightarrow[\rho \to 0]{} 0$$

$+$ non asymptotic convergence bounds when $\phi_\rho =$ Gaussian

Non-asymptotic guarantees for Gaussian smoothing

$$\pi_\rho(\mathbf{x}|\mathbf{y}) \propto \prod_{i=1}^{b} \int_{\mathbb{R}^{d_i}} \underbrace{e^{-f_i(\mathbf{z}_i)}}_{\pi_i(\mathbf{z}_i)} \underbrace{\mathcal{N}\left(\mathbf{z}_i | \mathbf{A}_i\mathbf{x}, \rho^2 \mathbf{I}_d\right)}_{\phi_\rho(\mathbf{z}_i, \mathbf{A}_i\mathbf{x})} \mathrm{d}\mathbf{z}_i$$

Distance	**Upper bound**	**Main assumptions**
	$\rho \sum_{i=1}^{b} 2\sqrt{d_i} L_i + o(\rho)$	f_i L_i-Lipschitz
$\\|\pi_\rho - \pi\\|_{\mathrm{TV}}$	$\frac{1}{2}\rho^2 M_1 d$	f_1 M_1-smooth, $b = 1$
	$\frac{1}{2}\rho^2 \sum_{i=1}^{b} M_i d_i + o(\rho^2)$	f_i M_i-smooth & strongly convex
$W_1(\pi_\rho, \pi)$	$\min(\rho\sqrt{d}, \frac{1}{2}\rho^2\sqrt{M_1 d})$	f_1 M_1-smooth, strongly convex

Splitted Gibbs sampling (SP): conditional distributions

Full conditional distributions under the split distribution π_ρ:

$$\pi_\rho(\mathbf{x}|\mathbf{z}) \propto \exp\left(-f_1(\mathbf{x}) - \phi_\rho(\mathbf{x}, \mathbf{z})\right)$$

$$\pi_\rho(\mathbf{z}|\mathbf{x}) \propto \exp\left(-f_2(\mathbf{z}) - \phi_\rho(\mathbf{x}, \mathbf{z})\right).$$

Note that f_1 and f_2 are now separated in **2 distinct distributions**

State of the art sampling methods:

- Gaussian variables: Fourier or Aux-V1 or E-PO
- MYULA = proximal MCMC,
- ...

Splitted Gibbs sampling (SP): conditional distributions

Full conditional distributions under the split distribution π_ρ:

$$\pi_\rho(\mathbf{x}|\mathbf{z}) \propto \exp\left(-f_1(\mathbf{x}) - \frac{1}{2\rho^2}\|\mathbf{x}-\mathbf{z}\|_2^2\right)$$

$$\pi_\rho(\mathbf{z}|\mathbf{x}) \propto \exp\left(-f_2(\mathbf{z}) - \frac{1}{2\rho^2}\|\mathbf{x}-\mathbf{z}\|_2^2\right).$$

Note that f_1 and f_2 are now separated in **2 distinct distributions**

State of the art sampling methods:

- Gaussian variables: Fourier or Aux-V1 or E-PO
- MYULA = proximal MCMC,
- ...

Partial conclusion

Efficient sampling for inverse problems in high dimensions

- **Inverse problems and Bayesian inference**
  - optimization
  - the usual Bayesian toolbox
  - focus on Langevin sampling: **ULA**, **MALA**, **MYULA**
- **SGS & SPA split-and-augment strategy**
  - Bayesian inference for complex models
  - large scale problems (big & tall)
  - **confidence intervals**
- **Efficient algorithms** for inference:
  - **acceleration** of state-of-the-art sampling algorithms
  - **distributed** inference (privacy, distr. comput.)
- **AXDA**: **unifying** statistical framework
  - asymptotically exact: control parameter ρ
  - **non-asymptotic theoretical guarantees**

Coming next: applications & extensions

- **Distributed sampling: fast and scalable**: **SPMD**
  - localized operators
  - distributed computing: coding
  - **confidence intervals**
- **unifying** statistical framework: **AXDA**
  - ELSA for PCGS: Mehdi Amrouche's PhD (J. Idier & H. Carfantan)
  - VAE prior + AXDA: Mario Gonzalez's PhD (A. Almansa, P. Muse)
- **Generative models** for inference: PnP-ULA & **PnP-SGS**
  - **learning** sampling networks
  - **evaluating** posterior distributions

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n},$$

where $\mathbf{A}$ = damaging operator and $\mathbf{n} \sim \mathcal{N}\left(\mathbf{0}_d, \sigma^2 \mathbf{I}_d\right)$ = noise.

$$\begin{cases} f_1(\mathbf{x}) & = \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \quad \forall \mathbf{x} \in \mathbb{R}^d, \\ f_2(\mathbf{x}) & = \tau \psi(\mathbf{x}), \quad \tau > 0. \end{cases}$$

Then the SP conditional distributions are:

$$\pi_\rho(\mathbf{x}|\mathbf{z}) = \mathcal{N}\left(\boldsymbol{\mu}_\mathbf{x}, \mathbf{Q}_\mathbf{x}{}^{-1}\right)$$
$$\pi_\rho(\mathbf{z}|\mathbf{x}) \propto \exp\left(-\tau\psi(\mathbf{z}) - \frac{1}{2\rho^2} \|\mathbf{z} - \mathbf{x}\|_2^2\right),$$

Splitted Gibbs sampling (SP/SGS): efficient sampling

Linear Gaussian inverse problems

$$\pi_\rho(\mathbf{x}|\mathbf{z}) = \mathcal{N}\left(\boldsymbol{\mu}_\mathbf{x}, {\mathbf{Q}_\mathbf{x}}^{-1}\right)$$

$$\pi_\rho(\mathbf{z}|\mathbf{x}) \propto \exp\left(-\tau\psi(\mathbf{z}) - \frac{1}{2\rho^2}\left\|\mathbf{z}-\mathbf{x}\right\|_2^2\right),$$

Examples:

- Tikhonov regularization

  $\psi(\mathbf{z}) = \|\mathbf{Q}\mathbf{z}\|_2^2 \Rightarrow$ **Gaussian variables**

  (e.g. **P** or **Q** diagonalizable in Fourier$\rightarrow$ E-PO)

- Convex non-smooth

  $\psi =$ **TV**, ℓ_1 sparsity... $\Rightarrow$ **proximal MCMC**

Posterior distribution

$$p\left(\mathbf{x}|\mathbf{y}\right) \propto \exp\left[-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 - \beta\mathrm{TV}(\mathbf{x})\right]$$

where $\mathbf{P}$ = damaging operator (blur, binary mask...) and

$$\mathrm{TV}(\mathbf{x}) = \sum_{1\leq i,j\leq N} \left\|(\nabla\mathbf{x})_{i,j}\right\|_2$$

Direct sampling is challenging

1. generally **high dimension** of the image,
2. **non-conjugacy** of the TV-based prior,
3. **non-differentiability** of g ($\neq$ Hamiltonian Monte Carlo algorithms)

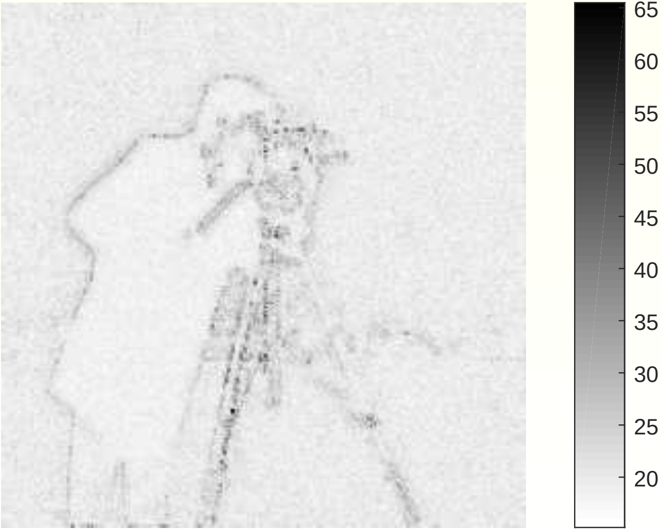
Splitted Gibbs sampling (SP/SGS): TV deblurring

Linear Gaussian inverse problems

Splitted Gibbs sampling (SP/SGS): TV deblurring

Splitted Gibbs sampling (SP): TV deblurring

Linear Gaussian inverse problems + **90% credibility intervals**



Splitted Gibbs sampling (SP/SGS): TV deblurring

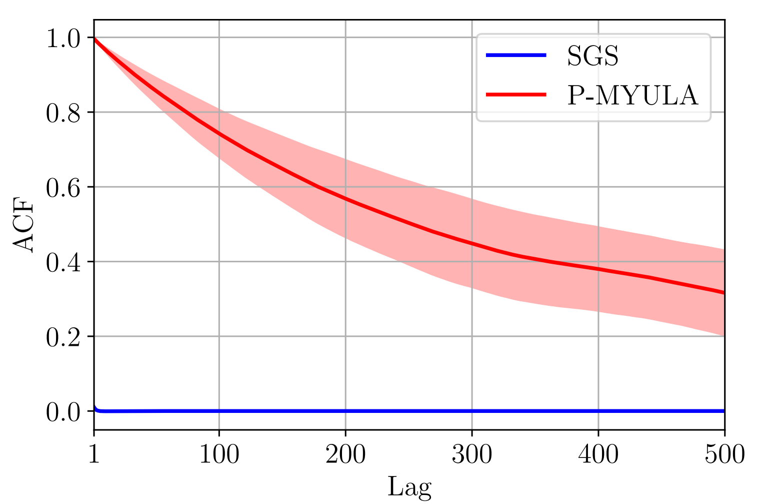
	SALSA	**FISTA**	**SGS**	**P-MYULA**
time (s)	1	10	470	3600
time ($\times$ var. split.)	1	10	**1**	**7.7**
nb. iterations	22	214	$\sim 10^4$	10^5
SNR (dB)	17.87	17.86	**18.36**	17.97

Rk : $\rho^2 = 9$

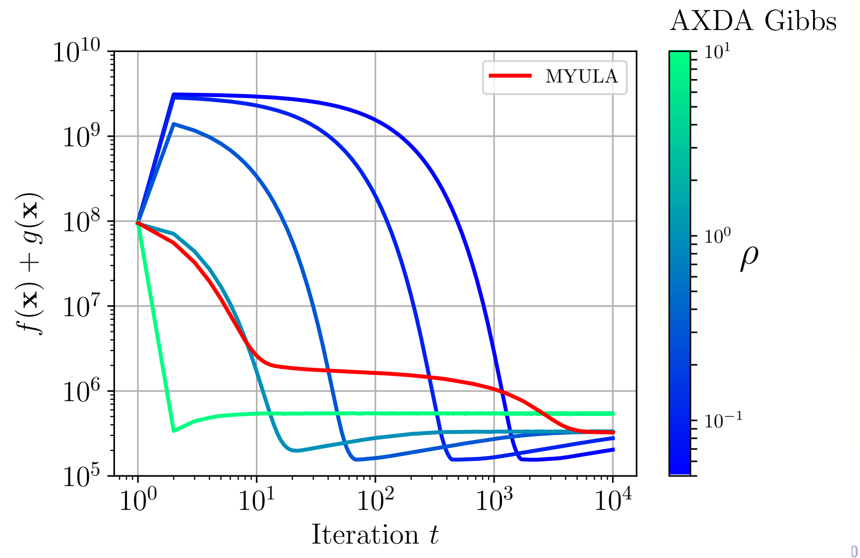
Splitted Gibbs sampling (SP): TV deblurring

Linear Gaussian inverse problems

Short auto-correlation of the Markov chain



$\rho =$ comput. time/quality trade-off



Splitted & Augmented Gibbs sampling (SPA) (optional)

Motivation for augmentation:

better mixing properties of the Markov chain

$$\begin{aligned}\pi_{\rho,\alpha} &\triangleq p(\mathbf{x}, \mathbf{z}, \mathbf{u}; \rho, \alpha) \\ &\propto \exp\left[-f(\mathbf{x}) - g(\mathbf{z})\right] \\ &\times \exp\left[-\phi_1(\mathbf{x}, \mathbf{z} - \mathbf{u}; \rho) - \phi_2(\mathbf{u}; \alpha)\right]\end{aligned}$$

Assumption 2

ϕ_2 and ϕ_1 are such that $\forall \mathbf{x}, \mathbf{z} \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} \exp\left[-\phi_1(\mathbf{x}, \mathbf{z} - \mathbf{u}; \rho) - \phi_2(\mathbf{u}; \alpha)\right] \mathrm{d}\mathbf{u} \propto \exp\left[-\phi_1(\mathbf{x}, \mathbf{z}; \eta(\rho, \alpha))\right].$$

The conditional split-augmented distributions are:

$$p(\mathbf{x}|\mathbf{z},\mathbf{u};\rho) \propto \exp\left[-f(\mathbf{x}) - \frac{1}{2\rho^2}\|\mathbf{x}-\mathbf{z}+\mathbf{u}\|_2^2\right]$$

$$p(\mathbf{z}|\mathbf{x},\mathbf{u};\rho) \propto \exp\left[-g(\mathbf{z}) - \frac{1}{2\rho^2}\|\mathbf{x}-\mathbf{z}+\mathbf{u}\|_2^2\right]$$

$$p(\mathbf{u}|\mathbf{x},\mathbf{z};\rho,\alpha) \propto \exp\left[-\frac{\|\mathbf{u}\|_2^2}{2\alpha^2} - \frac{1}{2\rho^2}\|\mathbf{x}-\mathbf{z}+\mathbf{u}\|_2^2\right].$$

By replacing Gibbs sampling steps by optimizations, ADMM appears:

$$\mathbf{x}^{(t)} \in \arg\min_{\mathbf{x}} -\log p\left(\mathbf{x}|\mathbf{z}^{(t-1)}, \mathbf{u}^{(t-1)}; \rho\right)$$

$$\mathbf{z}^{(t)} \in \arg\min_{\mathbf{z}} -\log p\left(\mathbf{z}|\mathbf{x}^{(t)}, \mathbf{u}^{(t-1)}; \rho\right)$$

$$\mathbf{u}^{(t)} = \mathbf{u}^{(t-1)} + \mathbf{x}^{(t)} - \mathbf{z}^{(t)}$$

Outline

Many problems can be considered using AXDA/SPA:

- Laplacian + ℓ_2 regularizer for deconvolution

  M. Vono et al., "Split-and-augmented Gibbs sampler - Application to large-scale inference problems," in *IEEE Trans. Signal Processing*, 2019

- Poisson noise + blur + non-negativity + ...

  M. Vono et al., "Bayesian image restoration under Poisson noise and log-concave prior," in *Proc. ICASSP 2019*

- Machine learning: logistic regression,...

  M. Vono et al. (2018), "Sparse Bayesian binary logistic regression using the split-and-augmented Gibbs sampler," in *Proc. IEEE MLSP 2018*

Splitted & Augmented Gibbs sampling (SGS) in action

Applications

Many problems can be considered using AXDA/SPA:

- Laplacian + ℓ_2 regularizer for deconvolution

  M. Vono et al., "Split-and-augmented Gibbs sampler - Application to large-scale inference problems," in *IEEE Trans. Signal Processing*, 2019

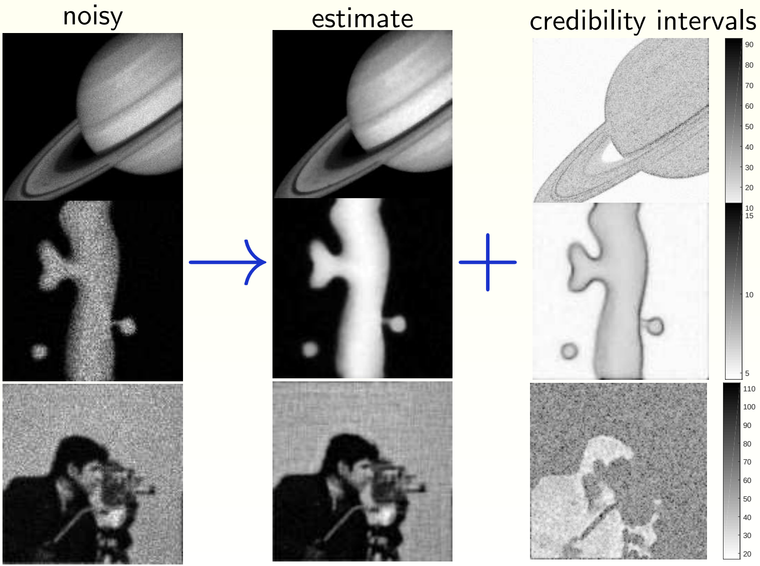
- Poisson noise + blur + non-negativity + ...

  M. Vono et al., "Bayesian image restoration under Poisson noise and log-concave prior," in *Proc. ICASSP 2019*

- Machine learning: logistic regression,...

  M. Vono et al. (2018), "Sparse Bayesian binary logistic regression using the split-and-augmented Gibbs sampler," in *Proc. IEEE MLSP 2018*

Bayesian image restoration under Poisson noise



Vono et al. (2019)

Take-home message

- **Motivations**
  - Posterior distr. of estimators $\rightarrow$ **Bayesian** + **MCMC**
  - Quantify **uncertainty**
- **Challenges**
  - **Poisson** likelihood and **distributed data**
  - Sophisticated prior $\rightarrow$ **difficult sampling**
- **Contributions**
  - **Variable splitting** for MCMC (akin to the ADMM)
  - **Fast**, **general** MCMC strategy
  - **State-of-the-art** performance

Problem statement

Model: $\forall n \in [\![1, N]\!]$,

$$y_n \sim \mathrm{Poisson}([\mathbf{A}\mathbf{x}]_n), \quad \mathbf{A} = \text{blurring operator.}$$

Neg. log likelihood:

$$\sum_{n=1}^{N} -y_n \log\left([\mathbf{A}\mathbf{x}]_n\right) + [\mathbf{A}\mathbf{x}]_n.$$

Prior:

- $\mathbf{x} \succeq \mathbf{0}_d$
- Total variation, ℓ_1, ...

Solution: splitting!

General problem formulation

Let $\pi \in L^1$ be the target posterior with neg. log density

$$-\log \pi(\mathbf{x}) = \underbrace{\sum_{n=1}^{N} f_n(\mathbf{A}_n \mathbf{x})}_{\text{neg. log likelihood}} + \underbrace{\sum_{k=1}^{K} g_k(\mathbf{G}_k \mathbf{x})}_{\text{neg. log prior}}.$$

- $\mathbf{A}_n$: blur, binary mask, ...
- f_n: Poisson neg. log likelihood.
- $\mathbf{G}_k$: transform, dictionary, ...
- g_k: ℓ_1 norm, TV regularization, $\iota_{\mathbb{R}_+^d}$, ...

General problem formulation

Let $\pi \in L^1$ be the target posterior with neg. log density

$$-\log \pi(\mathbf{x}) = \underbrace{\sum_{n=1}^{N} f_n(\mathbf{A}_n \mathbf{x})}_{\text{neg. log likelihood}} + \underbrace{\sum_{k=1}^{K} g_k(\mathbf{G}_k \mathbf{x})}_{\text{neg. log prior}}.$$

Issues:

- f_n continuously differentiable but not grad. Lipschitz
- large N + distributed data
- non-conjugate + non-smooth + multi-potential prior.

Variable splitting and the ADMM

Take inspiration from the **ADMM** and its **variable-splitting** formulation.

Reminder:

$$\min_{\mathbf{x}} \; f_1(\mathbf{A}_1\mathbf{x}) + g_1(\mathbf{G}_1\mathbf{x})$$

becomes

$$\min_{\mathbf{x},\mathbf{z}_1,\mathbf{u}_1} \; f_1(\mathbf{u}_1) + g_1(\mathbf{z}_1)$$

such that

$$\mathbf{u}_1 = \mathbf{A}_1\mathbf{x} \text{ and } \mathbf{z}_1 = \mathbf{G}_1\mathbf{x}$$

Variable splitting for sampling

Neg. log. posterior (cost function):

$$-\log \pi(\mathbf{x}) = \underbrace{\sum_{n=1}^{N} f_n(\mathbf{A}_n \mathbf{x})}_{\text{neg. log likelihood}} + \underbrace{\sum_{k=1}^{K} g_k(\mathbf{G}_k \mathbf{x})}_{\text{neg. log prior}}.$$

Variable splitting → joint distribution $p_\rho(\mathbf{x}, \mathbf{z}_{1:K}, \mathbf{u}_{1:N})$ such that

$$-\log p_\rho(\mathbf{x}, \mathbf{z}_{1:K}, \mathbf{u}_{1:N}) = \underbrace{\sum_{n=1}^{N} f_n(\mathbf{u}_n)}_{\text{split likelihood}} + \underbrace{\sum_{k=1}^{K} g_k(\mathbf{z}_k)}_{\text{split prior}}$$
$$+ \sum_{n=1}^{N} \frac{1}{2\rho^2} \left\| \mathbf{u}_n - \mathbf{A}_n \mathbf{x} \right\|^2 + \sum_{k=1}^{K} \frac{1}{2\rho^2} \left\| \mathbf{z}_k - \mathbf{G}_k \mathbf{x} \right\|^2 .$$

Split Gibbs sampler (SGS)

Sample from $p_\rho(\mathbf{x}, \mathbf{z}_{1:K}, \mathbf{u}_{1:N})$ with a **simple**, **efficient** and **theoretically sound** Gibbs sampler:

$$p_\rho(\mathbf{u}_n|\mathbf{x}) \propto \exp\left(-f_n(\mathbf{u}_n) - \frac{1}{2\rho^2}\|\mathbf{u}_n - \mathbf{A}_n\mathbf{x}\|^2\right)$$

$$p_\rho(\mathbf{z}_k|\mathbf{x}) \propto \exp\left(-g_k(\mathbf{z}_k) - \frac{1}{2\rho^2}\|\mathbf{z}_k - \mathbf{G}_k\mathbf{x}\|^2\right)$$

$p_\rho(\mathbf{x}|\mathbf{z}_{1:K}, \mathbf{u}_{1:N})$: d-dimensional Gaussian.

Sampling: P-MYULA (Durmus et al. 2018b) for auxiliary $\mathbf{u}_n$, $\mathbf{z}_k$
E-PO (Papandreou and Yuille 2010) for Gaussian $\mathbf{x}$.

Highlight

High-dimensional Gaussian sampling:
a review and a unifying approach based on
a stochastic proximal point algorithm

M. Vono, N. Dobigeon and P. C.

SIAM Review, vol. 64, no. 1, pp. 3-56, 2022

Vono et al. (2022a)

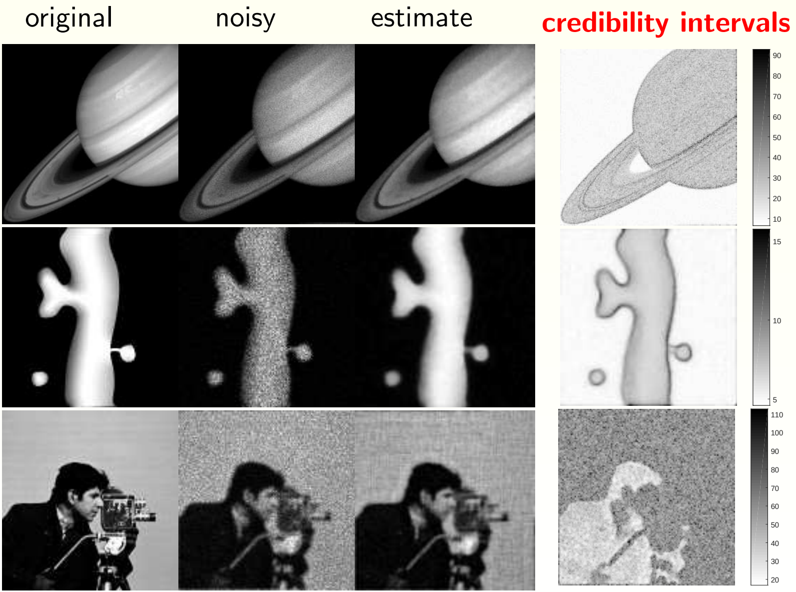
Experimental design

Model: $\forall n \in [\![1, N]\!]$, $y_n \sim \mathrm{Poisson}([\mathbf{A}x]_n)$, $\mathbf{A}$: blurring operator.

Prior: $g_1 = \iota_{\mathbb{R}_+^d}$,
g_2: $\tau\ell_1$ or $\tau\mathrm{TV}$, $\tau > 0$.

Images: 3 standard images with different intensity levels M.

Results


original
noisy
estimate
credibility intervals
90
80
70
60
50
40
30
20
10
15
10
5
110
100
90
80
70
60
50
40
30
20

Results

norm. MAE = mean absolute error / M.

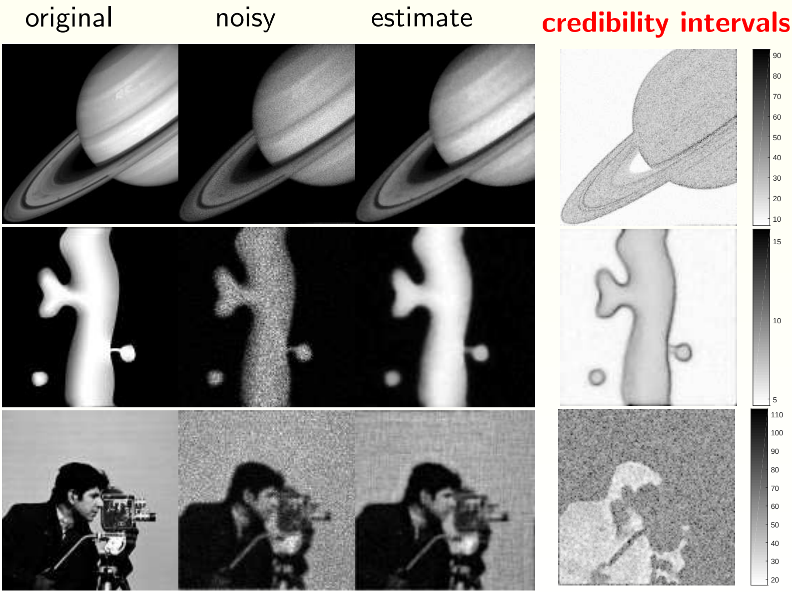
				norm. MAE		
image	**approach**	M	τ	**PIDAL**	**P-MYULA**	**SGS**
Saturn	TV	300	0.1	0.01	0.01	0.01
neuron	TV	30	1	0.03	0.03	0.05
		100	1	0.03	0.03	0.03
cameraman	WT	30	0.1	0.08	0.07	0.10
		100	0.1	0.07	0.06	0.07
		255	0.1	0.07	0.06	0.06

State-of-the-art performance with **controlled approximations**.

Speed: SGS is **7 times faster** than state-of-the-art MCMC,

only $\sim$ **40-100 times slower** than ADMM.

Results


original
noisy
estimate
credibility intervals
90 80 70 60 50 40 30 20 10
15 10 5
110 100 90 80 70 60 50 40 30 20

- **efficient & simple** MCMC splitting strategy
  - divide-and-conquer
  - embeds & accelerate state-of-the-art algorithms
  - yields comprehensive and excellent results.

- based on the AXDA **unifying** statistical framework
  - mixture-based models
  - robust Bayesian models
  - variable splitting-based models

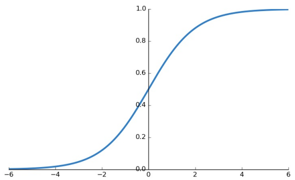
- **non-asymptotic** theoretical guarantees on the approximation under mild assumptions + **explicit convergence rates**.

Distributed sampling and data privacy

Regularized logistic regression by applying AXDA b times

$$\forall\ i \in [\![1, n]\!], \quad y_i \sim \text{Bernoulli}\left(\sigma(\mathbf{a}_i{}^T \mathbf{x})\right)$$

$$\pi(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\sum_{j=1}^{b} g^{(j)}(\mathbf{x}) - f(\mathbf{x})\right)$$



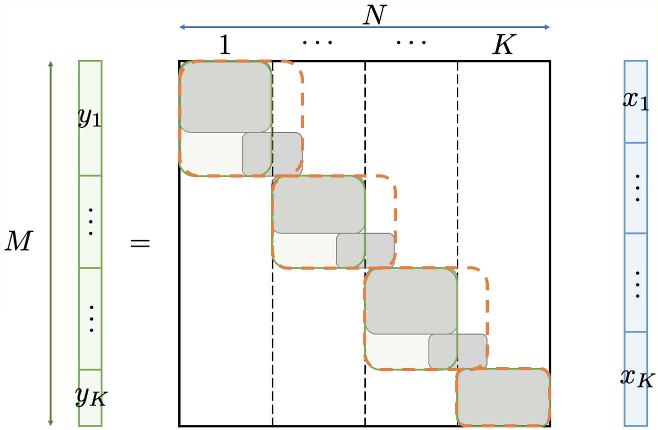
- $g^{(j)}(\mathbf{x}) = \sum_{i \in \mathcal{D}_j} \log\left(1 + \exp\left(-y_i \mathbf{a}_i{}^T \mathbf{x}\right)\right)$,
- $\mathcal{D}_j$ indices of the jth block of data,
- f = prior on the regressor $\mathbf{x}$

- inference via a Gibbs sampler distributed on b nodes
- the master node never *sees* the data set: **privacy**
- theoretical guarantees on the approximation

$$p_\rho(\mathbf{x}, \mathbf{z}_{1:b}) \propto \exp\left(-\sum_{j=1}^{b}\left[\frac{1}{2\rho^2}\|\mathbf{x} - \mathbf{z}_j\|^2 + \sum_{i \in \mathcal{D}_j} \log\left(1 + \exp\left(-y_i \mathbf{a}_i^T \mathbf{z}_j\right)\right)\right] - f(\mathbf{x})\right)$$

Fast distributed sampling: leveraging many CPUs

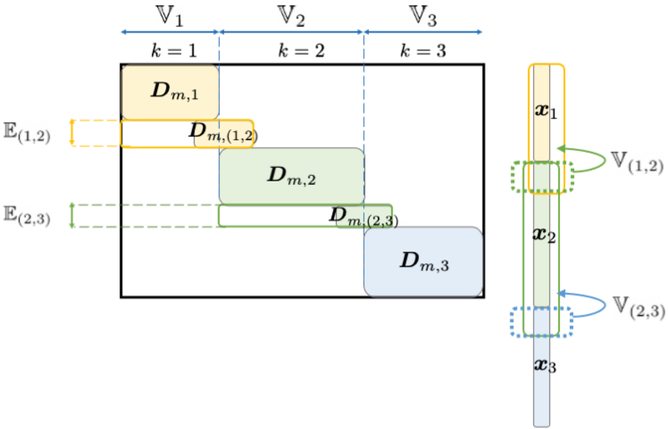
Collab. **P.-A. Thouvenin** & **A. Repetti** (Edinburgh)
Particular case : **localized** observation operators $\mathbf{A}$



Thouvenin et al. (2022a,b), preprint arXiv

Fast distributed sampling: leveraging many CPUs

Collab. **P.-A. Thouvenin** & **A. Repetti** (Edinburgh)
Particular case : **localized** observation operators **A**

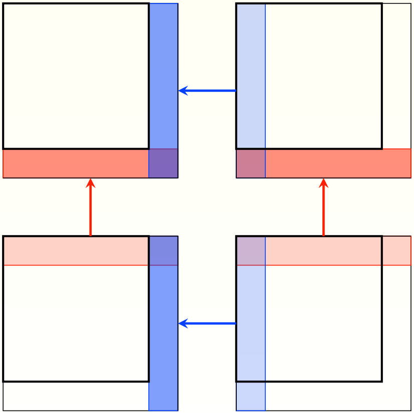


Thouvenin et al. (2022a,b), preprint arXiv

Fast distributed sampling: leveraging many CPUs

Collab. **P.-A. Thouvenin** & **A. Repetti** (Edinburgh)

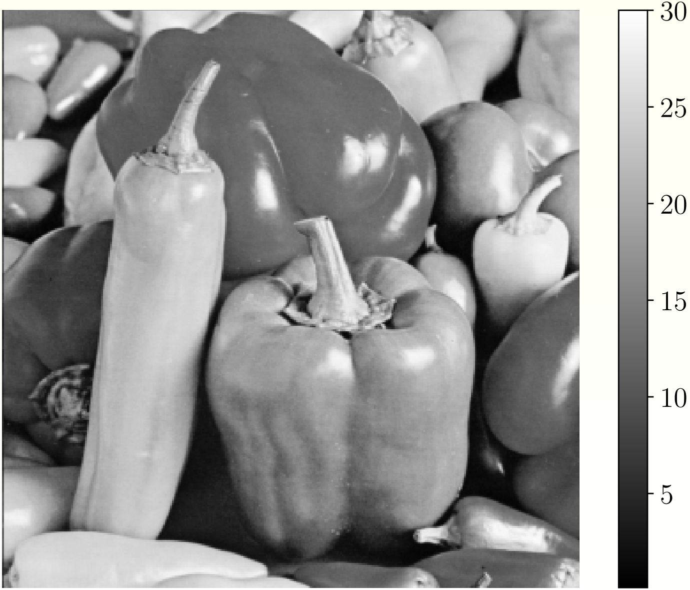
Particular case : **localized** observation operators **A**



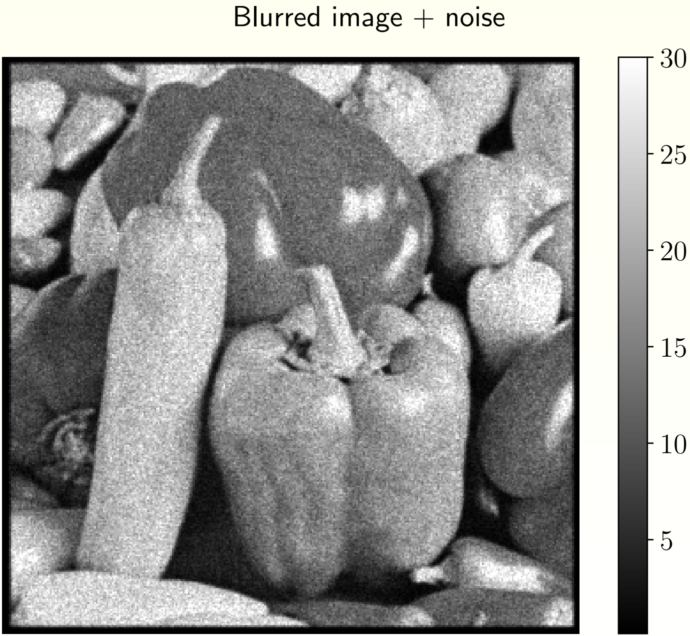
Thouvenin et al. (2022a,b), preprint arXiv

Fast distributed sampling: leveraging many CPUs

Original image



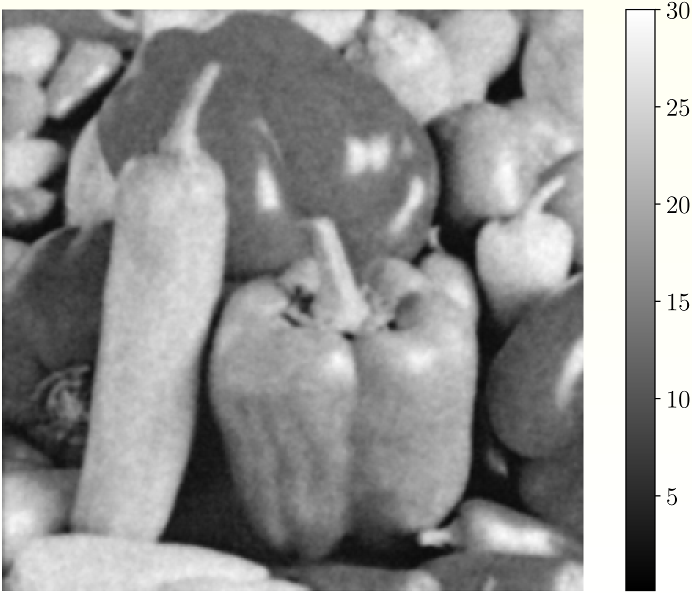
Fast distributed sampling: leveraging many CPUs



Fast distributed sampling: leveraging many CPUs

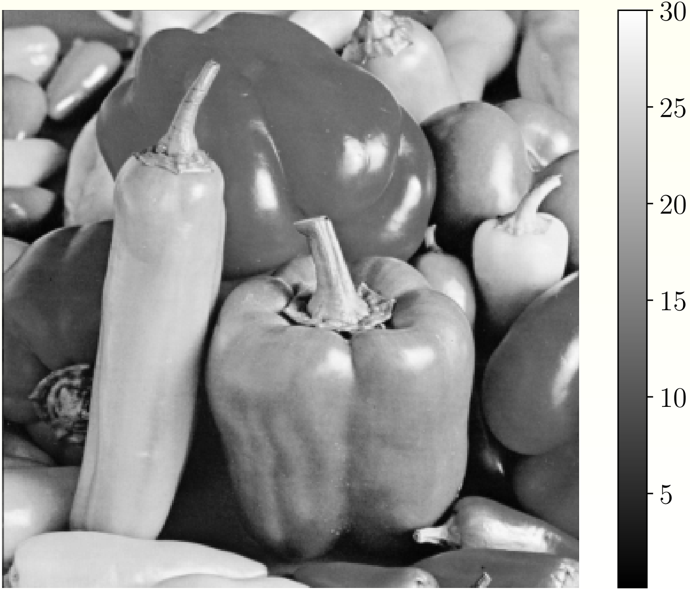
Fast distributed sampling: leveraging many CPUs

Dist. inference



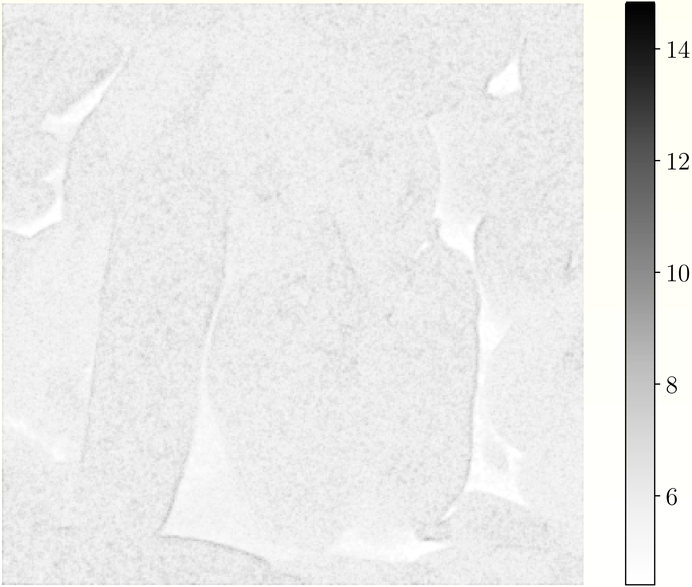
Fast distributed sampling: leveraging many CPUs

Original image



Fast distributed sampling: leveraging many CPUs

Credibility intervals



Fast distributed sampling: leveraging many CPUs

Particular case : **localized** observation operators $\mathbf{A}$

Collab. **P.-A. Thouvenin** & **A. Repetti** (Edinburgh)

- splitting the global variable of interest into blocks
- distributed block-coordinate SPA-Gibbs sampler

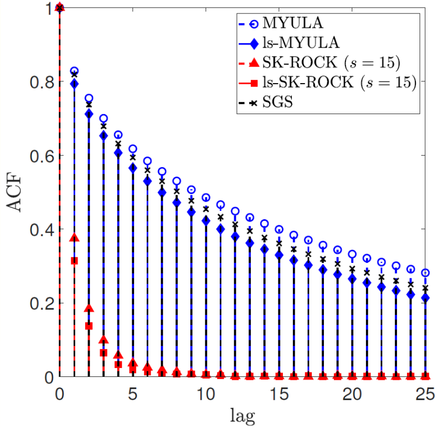
Method (cores)	SGS(1)	Dist.(1)	Dist.(2)	Dist.(16)
ms/sample	65.56	12.21	6.07	**1.08**
Accel. factor	0.19	1	2.01	**11.30**
Total time (s)	262.20	61.04	30.37	**5.38**
SNR (dB)	23.33	23.45	23.46	**23.48**

Related works

- Distributed MCMC
  Rendell et al. (2021); Plassier et al. (2021)
- M. Gonzalez et al.: Joint posterior MAP and posterior sampling with VAE prior
  González et al. (2022)
- M. Amrouche et al.: ELSA for partially collapsed Gibbs sampling (PCGS): asymptotically Exact Location Scale mixture Approximation for Bernoulli-$\mathcal{D}$ problems
  Amrouche et al. (2022)
- theoretical guarantees
  Durmus and Moulines (2017); Vono et al. (2022b); Laumont et al. (2022)
- L. Vargas, M. Pereyra et al.: Accelerated sampling using Runge-Kutta discretization of Langevin equation: SK-ROCK + SGS - Pereyra et al. (2020)

Related works

- L. Vargas, M. Pereyra et al., Accelerated sampling using Runge-Kutta discretization of Langevin stochastic equation $\Rightarrow$ SK-ROCK + SGS



[courtesy Pereyra et al. (2020), inpainting problem]

Outline

Capitalizing on machine learning

Various possible approaches

- **Deep learning**
  - strong expressivity
  - very efficient for supervised learning
  - large data set needed
- **Direct inversion**: $\mathbf{y} \Rightarrow \mathbf{x}$
  - fully supervised setting
  - Dong et al. (2014); Gao et al. (2019); Schwartz et al. (2018)
- **Deep image prior**
  - choose an architecture for $\mathbf{x} = f_\theta(\mathbf{z})$, fix $\mathbf{z}$ & optimize θ
  - similar to sparsity assumptions (functional analysis)
  - Ulyanov et al. (2018)
- **Learnt priors**
  - generative models: learn $\mathbf{x} = f_\theta(\mathbf{z})$; known $p(\mathbf{z}) \Rightarrow p(\mathbf{x})$
  - $\text{prox}_{f_2} \simeq$ denoisers : **Plug-and-Play (PnP) approaches**
  - Venkatakrishnan et al. (2013); Zhang et al. (2021)

- **Replacing prox_{f_2} by** a trained MAP denoiser

Recall (the zoo of prox):

$$\mathrm{prox}_{\lambda f_2}(\mathbf{x}) = \arg\min_{\mathbf{z}} \lambda f_2(\mathbf{z}) + \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2$$

Proximal operator = denoiser with prior $\propto \exp[-f_2(\mathbf{x})]$:

$$D_\varepsilon^\dagger(\mathbf{x}) = \arg\min_{\mathbf{z}} \varepsilon f_2(\mathbf{z}) + \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2$$

- **PnP-ADMM** $\Longleftarrow$ replace $\mathrm{prox}_{\lambda f_2}(\mathbf{x})$ by $D_\varepsilon^\dagger(\mathbf{x})$
- Chan et al. (2016)

Rk: any denoiser may not correspond to some prox_{f_2}; pb of theoretical guarantees...

Capitalizing on machine learning

PnP & gradient descent: **PnP-ULA**

- Using an MMSE denoiser & Tweedie's identity

$$D_\varepsilon^*(\mathbf{x}) = \mathbf{E}[\mathbf{x}|\mathbf{x}_\varepsilon]$$

$p_\varepsilon(\mathbf{x}) = p * \mathcal{N}(\cdot; \mathbf{x}, \varepsilon) \Rightarrow$ Tweedie's identity:

$$-\nabla \log p_\varepsilon(\mathbf{x}) = \frac{1}{\varepsilon}[x - D_\varepsilon^*(\mathbf{x})]$$

From **MYULA**

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \delta \nabla f_1(\mathbf{x}) + \delta \frac{1}{\lambda}[\mathrm{prox}_{\lambda f_2}(\mathbf{x}) - \mathbf{x}] + \sqrt{2\delta}\, \mathbf{w}^{(k+1)}$$

to **PnP-ULA**

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} \underbrace{- \delta \nabla f_1(\mathbf{x})}_{\textit{likelihood}} + \delta \underbrace{\frac{1}{\varepsilon}[D_\varepsilon^*(\mathbf{x}) - \mathbf{x}]}_{\textit{prior}} + \sqrt{2\delta}\, \mathbf{w}^{(k+1)}$$

Laumont et al. (2022)

- **PnP-SGS**: using a deep denoiser as a prior in SGS

SGS uses Gibbs sampling from conditional posteriors

$$p(\mathbf{x} \mid \mathbf{y}, \mathbf{z}) \quad \propto \quad \exp\left[-f(\mathbf{y}, \mathbf{x}) - \frac{1}{2\rho^2}\|\mathbf{x} - \mathbf{z}\|_2^2\right] \tag{1}$$

$$p(\mathbf{z} \mid \mathbf{x}) \quad \propto \quad \exp\left[-g(\mathbf{z}) - \frac{1}{2\rho^2}\|\mathbf{x} - \mathbf{z}\|_2^2\right] \tag{2}$$

- **PnP-SGS**: using a deep denoiser as a prior in SGS

DDPM: Denoising Diffusion Probabilistic Models

Learn backward SDE denoiser

$$p_{\boldsymbol{\theta}}\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t\right) = \mathcal{N}\left(\mathbf{x}_{t-1}; \boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\mathbf{x}_t, t\right), \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\left(\mathbf{x}_t, t\right)\right)$$

$$\Rightarrow p\left(\mathbf{z} \mid \mathbf{x}\right)$$

Trained from forward SDE "noising"

$$q\left(\mathbf{x}_t \mid \mathbf{x}_{t-1}\right) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{1 - b(t)}\mathbf{x}_{t-1}, b(t)\mathbf{I}\right)$$

Capitalizing on machine learning

Plug-and-Play and splitting: **PnP-SGS** - F. Cœurdoux's PhD, N. Dobigeon - IRIT

- **PnP-SGS**: using a deep denoiser as a prior in SGS

Original image - noisy masked - PnP-ADMM - PnP-SGS - 90% cred. int.

- **PnP-SGS**: using a deep denoiser as a prior in SGS

Quantitative evaluation (FID, LPIPS, PSNR, SSIM) of solutions: inpainting 1000 images FFHQ 256×256. **Best**, <u>Second</u>.

	FID $\downarrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$
PnP-SGS	**38.36**	**0.144**	**23.59**	**0.813**
TV-SGS	71.12	0.785	21.09	0.524
PnP-ADMM	123.61	0.692	<u>22.41</u>	0.325
TV-ADMM	181.56	0.463	22.03	<u>0.784</u>
Score-SDE	76.54	0.612	13.52	0.437
DDRM	69.71	0.587	9.19	0.319
MCG	<u>39.26</u>	<u>0.286</u>	21.57	0.751

- **PnP-SGS**: using a deep denoiser as a prior in SGS

Runtime for each algorithm in Wall-clock time (computed with a single GTX 2080Ti GPU).

Method	Wall-clock time (s)	Ref.
Score-SDE	36.71	Song et al. (2022)
DDRM	2.03	Kawar et al. (2022)
MCG	80.10	Chung et al. (2023)
PnP-ADMM	3.63	Chan et al. (2016)
SGS-ULA	218.90	Vono et al. (2019b)
PnP-SGS	13.81	latest news

Capitalizing on machine learning

Plug-and-Play and splitting: **PnP-SGS** - F. Cœurdoux's PhD, N. Dobigeon - IRIT

original image

noisy masked image

PnP-SGS MMSE

Capitalizing on machine learning

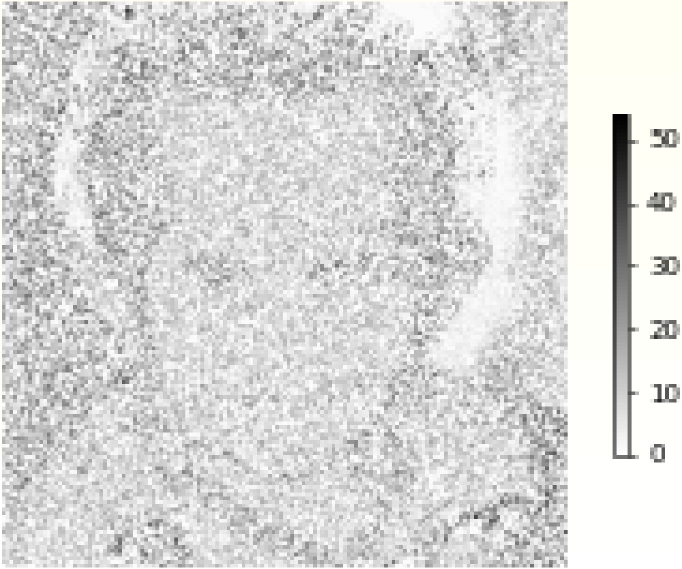
Plug-and-Play and splitting: **PnP-SGS** - F. Cœurdoux's PhD, N. Dobigeon - IRIT

original image

Capitalizing on machine learning

Plug-and-Play and splitting: **PnP-SGS** - F. Cœurdoux's PhD, N. Dobigeon - IRIT

90% credibility intervals (PnP-SGS)



original image

Capitalizing on machine learning

Plug-and-Play and splitting: **PnP-SGS** - F. Cœurdoux's PhD, N. Dobigeon - IRIT

noisy masked image

Capitalizing on machine learning

Plug-and-Play and splitting: **PnP-SGS** - F. Cœurdoux's PhD, N. Dobigeon - IRIT

PnP-SGS MMSE

Capitalizing on machine learning

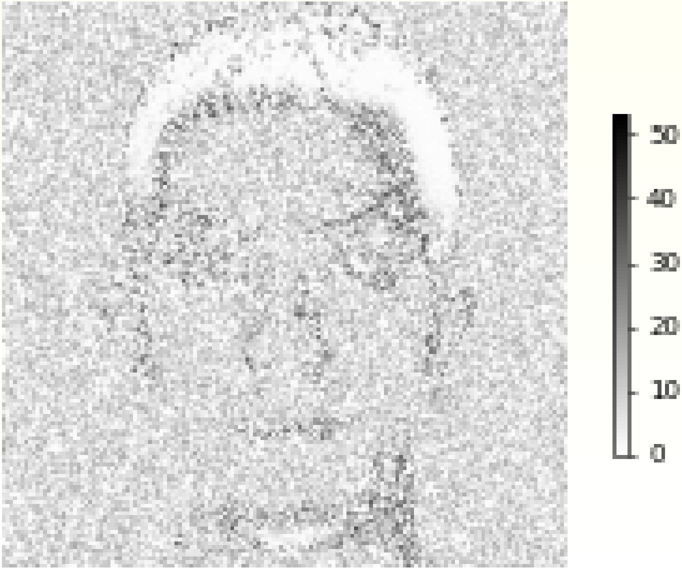
Plug-and-Play and splitting: **PnP-SGS** - F. Cœurdoux's PhD, N. Dobigeon - IRIT

original image

Capitalizing on machine learning

Plug-and-Play and splitting: **PnP-SGS** - F. Cœurdoux's PhD, N. Dobigeon - IRIT

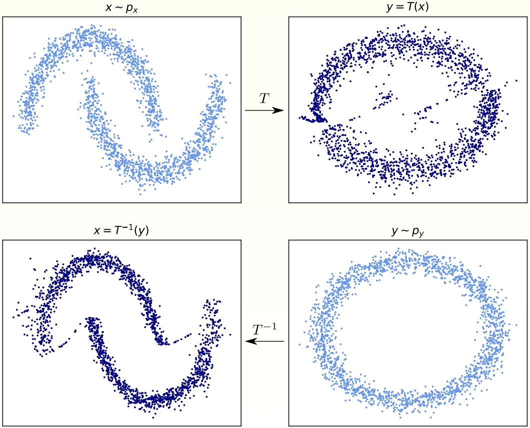
90% credibility intervals (PnP-SGS)



Related works: sampling using normalizing flows

F. Cœurdoux's PhD, N. Dobigeon - IRIT

(Deep) **Learning changes of variables** (optimal transport)



Coeurdoux et al. (2023) , preprint

Related works: sampling using normalizing flows

F. Cœurdoux's PhD, N. Dobigeon - IRIT

(Deep) **Learning changes of variables** (optimal transport)

$\Rightarrow$ combining MALA and Normalizing Flows...

- **MALAFlow**: sampling in the Gaussian latent space

Coeurdoux et al. (2023) , preprint

Outline

Efficient sampling for high dimensional problems



Nicolas Dobigeon, Maxime Vono, Pierre-Antoine Thouvenin



Pierre Palud, Audrey Repetti, Florentin Cœurdoux

Conclusion

Efficient sampling for inverse problems in high dimensions

- **SGS** & SPA split-and-augment strategy
  - Bayesian inference for complex models
  - large scale problems (big & tall)
  - **confidence intervals**
- **Efficient algorithms** for inference: ULA, MALA, MYULA
  - **acceleration** of state-of-the-art sampling algorithms
  - **distributed** inference (privacy, distr. comput.)
- **AXDA**: **unifying** statistical framework
  - asymptotically exact: control parameter ρ
  - **non-asymptotic theoretical guarantees**
- **Capitalizing on ML**: **trained denoisers**
  - learning from **representative samples**
  - theoretical guarantees under mild assumptions?

Applications & extensions

- **Distributed sampling: fast and scalable: SPMD**
  - localized operators
  - distributed computing: coding
  - **confidence intervals**
- **Generative models** for inference: PnP-ULA & **PnP-SGS**
  - **learning** sampling networks
  - **evaluating** posterior distributions
- **AXDA**: **unifying** statistical framework
  - ELSA for PCGS: Mehdi Amrouche's PhD (J. Idier & H. Carfantan)
  - VAE prior + AXDA: Mario Gonzalez's PhD (A. Almansa, P. Muse)

Interested in AXDA for your statistical problems?

- https://github.com/mvono

Amrouche, M., Carfantan, H., and Idier, J. (2022), "Efficient Sampling of Bernoulli-Gaussian-Mixtures for Sparse Signal Restoration," *IEEE Transactions on Signal Processing*, 70, 5578–5591.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011), "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends in Machine Learning*, 3, 1–122.

Chan, S. H., Wang, X., and Elgendy, O. A. (2016), "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Transactions on Computational Imaging*, 3, 84–98.

Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., and Ye, J. C. (2023), "Diffusion Posterior Sampling for General Noisy Inverse Problems," in *Proc. of Int. Conf. on Learning Representations*.

Coeurdoux, F., Dobigeon, N., and Chainais, P. (2023), "Learning Optimal Transport Between Two Empirical Distributions with Normalizing Flows," in *Machine Learning and Knowledge Discovery in Databases*, eds. Amini, M.-R., Canu, S., Fischer, A., Guns, T., Kralj Novak, P., and Tsoumakas, G., Cham: Springer Nature Switzerland, pp. 275–290.

Dong, C., Loy, C. C., He, K., and Tang, X. (2014), "Learning a deep convolutional network for image super-resolution," in *Proc. of Eur. Conf. on Comp. Vis.*, Springer, pp. 184–199.

Durmus, A. and Moulines, É. (2017), "Nonasymptotic convergence analysis for the unadjusted Langevin algorithm," *The Annals of Applied Probability*, 27, 1551 – 1587.

Durmus, A., Moulines, E., and Pereyra, M. (2018a), "Efficient Bayesian Computation by Proximal Markov Chain Monte Carlo: When Langevin Meets Moreau," *SIAM Journal on Imaging Sciences*, 11, 473–506.

— (2018b), "Efficient Bayesian Computation by Proximal Markov chain Monte Carlo: When Langevin Meets Moreau," *SIAM Journal on Imaging Sciences*, 11, 473–506.

Gao, H., Tao, X., Shen, X., and Jia, J. (2019), "Dynamic scene deblurring with parameter selective sharing and nested skip connections," in *Comput. Vision Pattern Recog.*, pp. 3848–3856.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003), *Bayesian Data Analysis*, Chapman and Hall/CRC, 2nd ed.

Geman, D. and Yang, C. (1995), "Nonlinear image recovery with half-quadratic regularization," *IEEE Transactions on Image Processing*, 4, 932–946.

González, M., Almansa, A., and Tan, P. (2022), "Solving Inverse Problems by Joint Posterior Maximization with Autoencoding Prior," *SIAM Journal on Imaging Sciences*, 15, 822–859.

Kawar, B., Elad, M., Ermon, S., and Song, J. (2022), "Denoising diffusion restoration models," *arXiv preprint arXiv:2201.11793*.

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, 701–737.

Neal, R. M. (2003), "Slice sampling," *The Annals of Statistics*, 31, 705–767.

Papandreou, G. and Yuille, A. L. (2010), "Gaussian sampling by local perturbations," in *Advances in Neural Information Processing Systems*, pp. 1858–1866.

Pereyra, M., Mieles, L. V., and Zygalakis, K. C. (2020), "Accelerating Proximal Markov Chain Monte Carlo by Using an Explicit Stabilized Method," *SIAM Journal on Imaging Sciences*, 13, 905–935.

Pereyra, M., Schniter, P., Chouzenoux, E., Pesquet, J.-C., Tourneret, J.-Y., Hero, A., and McLaughlin, S. (2016), "A Survey of Stochastic Simulation and Optimization Methods in Signal Processing," *IEEE Journal of Selected Topics in Signal Processing*, 10, 224–241.

Plassier, V., Vono, M., Durmus, A., and Moulines, E. (2021), "DG-LMC: a turn-key and scalable synchronous distributed MCMC algorithm via Langevin Monte Carlo within Gibbs," in *International Conference on Machine Learning (ICML)*.

Polson, N. G., Scott, J. G., and Windle, J. (2013), "Bayesian Inference for Logistic Models Using Polya-Gamma Latent Variables," *Journal of the American Statistical Association*, 108, 1339–1349.

Rendell, L. J., Johansen, A. M., Lee, A., and Whiteley, N. (2021), "Global consensus Monte Carlo," *Journal of Computational and Graphical Statistics*, 30.

Robert, C. P. (2001), *The Bayesian Choice: from decision-theoretic foundations to computational implementation*, New York: Springer, 2nd ed.

Robert, C. P. and Casella, G. (2004), *Monte Carlo Statistical Methods*, Berlin: Springer, 2nd ed.

Scheffé, H. (1947), "A useful convergence theorem for probability distributions," *The Annals of Mathematical Statistics*, 18, 434–438.

Schwartz, E., Giryes, R., and Bronstein, A. M. (2018), "DeepISP: Toward learning an end-to-end image processing pipeline," *IEEE Trans. on Image Processing*, 28, 912–923.

Song, Y., Shen, L., Xing, L., and Ermon, S. (2022), "Solving Inverse Problems in Medical Imaging with Score-Based Generative Models," in *Proc. of Int. Conf. on Learning Representations*.

Thouvenin, P., Repetti, A., and Chainais, P. (2022a), "A versatile distributed MCMC algorithm for large scale inverse problems," in *30th European Signal Processing Conference, EUSIPCO 2022, Belgrade, Serbia, August 29 - Sept. 2, 2022*, IEEE, pp. 2016–2020.

Thouvenin, P.-A., Repetti, A., and Chainais, P. (2022b), "A distributed Gibbs Sampler with Hypergraph Structure for High-Dimensional Inverse Problems," .

Ulyanov, D., Vedaldi, A., and Lempitsky, V. (2018), "Deep Image Prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

van Dyk, D. A. and Meng, X.-L. (2001), "The Art of Data Augmentation," *Journal of Computational and Graphical Statistics*, 10, 1–50.

Venkatakrishnan, S. V., Bouman, C. A., and Wohlberg, B. (2013), "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*, IEEE, pp. 945–948.

Vono, M., Dobigeon, N., and Chainais, P. (2019), "Bayesian Image Restoration under Poisson Noise and Log-concave Prior," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1712–1716.

Vono, M., Dobigeon, N., and Chainais, P. (2019a), "Split-and-augmented Gibbs sampler - Application to large-scale inference problems," *IEEE Transactions on Signal Processing*, 67, 1648–1661.

— (2019b), "Split-and-augmented Gibbs sampler—Application to large-scale inference problems," *IEEE Transactions on Signal Processing*, 67, 1648–1661.

— (2022a), "High-Dimensional Gaussian Sampling: A Review and a Unifying Approach Based on a Stochastic Proximal Point Algorithm," *SIAM Review*, 64, 3–56.

Vono, M., Paulin, D., and Doucet, A. (2022b), "Efficient MCMC Sampling with Dimension-Free Convergence Rate using ADMM-type Splitting," *Journal of Machine Learning Research*, 23, 1–69.

Zhang, K., Li, Y., Zuo, W., Zhang, L., Van Gool, L., and Timofte, R. (2021), "Plug-and-play image restoration with deep denoiser prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44, 6360–6376.