Efficient sampling through variable splitting: credibility intervals

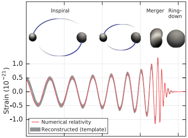
P. Chainais
P.-A. Thouvenin, P. Palud, C. Prévost,
A. Repetti (Heriot-Watt Univ., Edinburgh)
M. Vono, F. Cœurdoux, N. Dobigeon (IRIT)

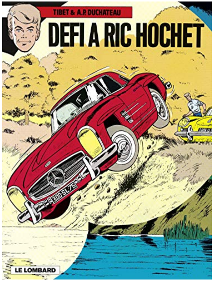
March 23rd 2022

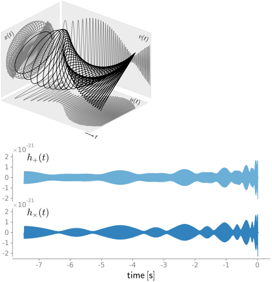
WP 2. Inverse problems and machine learning
for polarized signals reconstruction

Task 2.1. Bayesian models:
exploiting the geometry of the polarization information

Task 2.2. Machine learning for efficient Bayesian inference





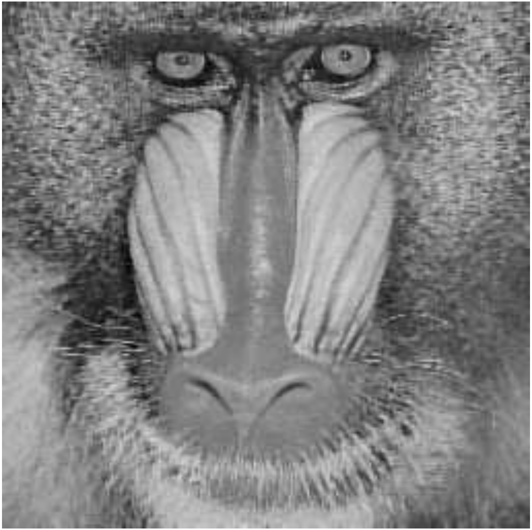


Motivations: applications in image processing

Image deblurring

Motivations: applications in image processing

Image deblurring



Motivations: applications in image processing

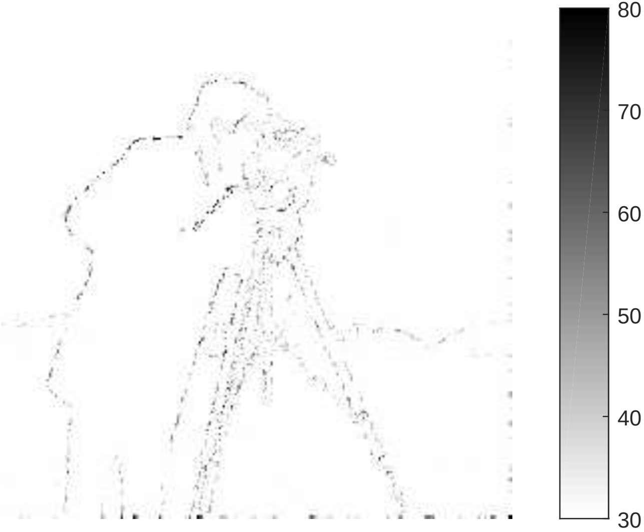
Image inpainting

Motivations: applications in image processing

Image inpainting

Motivations: applications in image processing

Confidence intervals



Motivations

- **solve complex** ill-posed inverse problems
- **big** data in **large** dimensions
- **excellent** performances
- **fast** inference algorithms
- **credibility intervals**

with maybe some additional options such as:

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

Motivations

- **solve complex** ill-posed inverse problems
- **big** data in **large** dimensions
- **excellent** performances
- **fast** inference algorithms
- **credibility intervals**

with maybe some additional options such as:

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

Bayesian approach + MCMC method (or even better?)

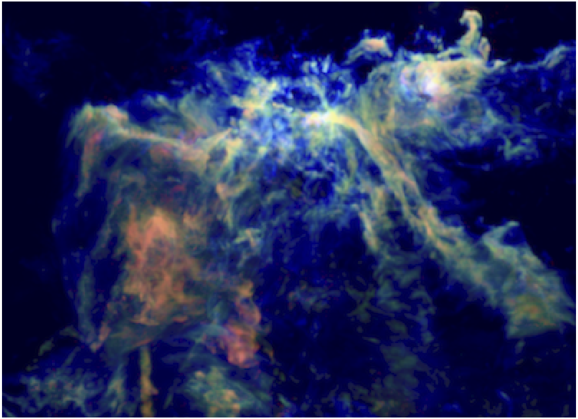
Motivations

The Bayesian approach: application to radio-astronomy - Orion-B

Astrophysics : absence of ground truth

Confidence intervals are crucial to acertain predictions

Pierre Palud's PhD with the Orion-B consortium



Motivations

The optimization-based approach

Inverse problems & **optimization**

= define a **cost function** : $f(\mathbf{x}) = f_1(\mathbf{x}, \mathbf{y}) + f_2(\mathbf{x})$

where f_2 is typically

- convex (or not)
- not differentiable $\Rightarrow$ proximal operators
- a sum of various penalties

Solution: proximal operators

Inverse problems & **optimization**

$$= \text{define a } \textbf{cost function} : \quad f(\mathbf{x}) = f_1(\mathbf{x}, \mathbf{y}) + f_2(\mathbf{x})$$

where f_2 is typically

- convex (or not)
- not differentiable $\Rightarrow$ proximal operators
- a sum of various penalties

Solution: **proximal operators** and **splitting** techniques

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

maybe relaxed to **ADMM**:

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

Motivations

The Bayesian approach

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

$$= \text{define a } \textbf{posterior distribution } p(\mathbf{x}|\mathbf{y}) \propto p_1(\mathbf{y}|\mathbf{x}) \cdot p_2(\mathbf{x})$$

where p_2 is typically

- log-concave (or not) $\leftrightarrow$ f_2 convex
- conjugate $\Rightarrow$ easy sampling/inference
- a combination of various prior

Solution: MCMC methods and **Gibbs** sampling

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

Splitting & Augmentation : **AXDA** and the **SPA sampler**

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

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

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

Computational motivations: difficult sampling

- non-conjugate priors [conj. priors $\Rightarrow$ easy inference]
- rich models: complicated prior distributions
- big datasets: expensive likelihood computation

Strategy: **Divide-To-Conquer**

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.

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)

One surrogate is to introduce auxiliary variables $\mathbf{z} \in \mathcal{Z} \subseteq \mathbb{R}^k$ 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** full 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

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. 2019)

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

- **simplify** the inference (Vono et al. 2019)
- **distribute** the inference (Rendell et al. 2018)
- **accelerate** the inference (Vono et al. 2019).

Splitted Gibbs sampling (SP)

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

$$\Downarrow$$

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

$$\Downarrow$$

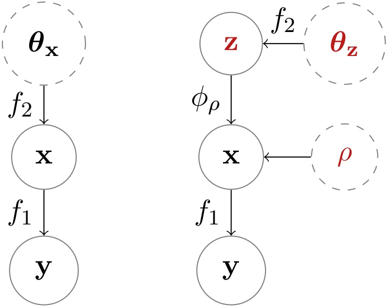
$$\pi_\rho(\mathbf{x}, \mathbf{z}) \propto \exp\left[-f_1(\mathbf{x}) - f_2(\mathbf{z}) - \frac{1}{2\rho^2}\|\mathbf{x} - \mathbf{z}\|_2^2\right]$$

Splitted Gibbs sampling (SP)

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

$$\Downarrow$$

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



Splitted Gibbs sampling (SP): Theorem 1

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

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

Splitted Gibbs sampling (SP): marginal 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**

Splitted Gibbs sampling (SP): marginal 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:

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

Splitted Gibbs sampling (SP): inverse problems

Linear Gaussian inverse problems

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

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

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

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): efficient sampling

Linear Gaussian inverse problems

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

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

Examples:

- Convex non-smooth

  $\psi = \mathbf{TV}$, $\ell_\mathbf{1}$ sparsity... $\Rightarrow$ **proximal MCMC**

- Tikhonov regularization

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

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

Posterior distribution

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

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): TV deblurring

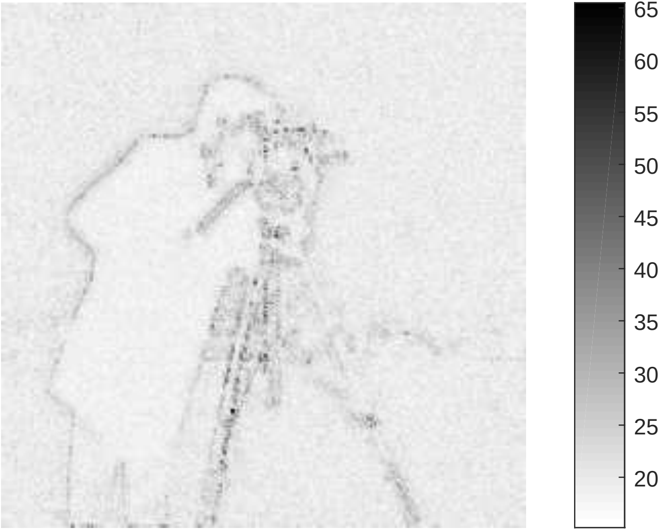
Linear Gaussian inverse problems

Splitted Gibbs sampling (SP): TV deblurring

Linear Gaussian inverse problems

Splitted Gibbs sampling (SP): TV deblurring

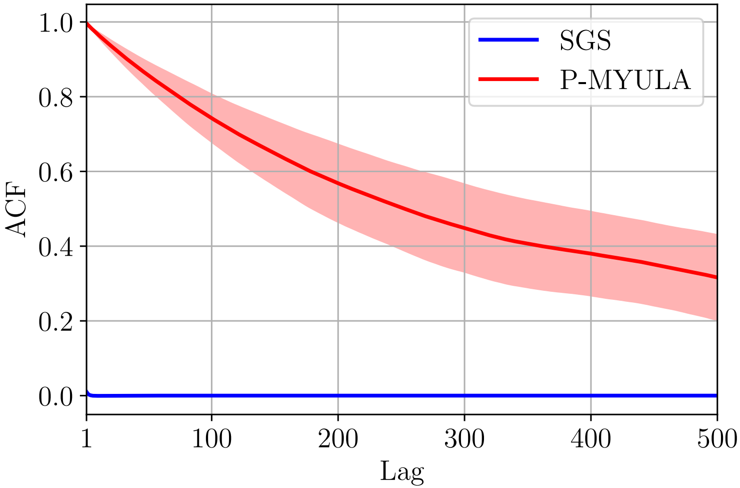
Linear Gaussian inverse problems + **90% credibility intervals**



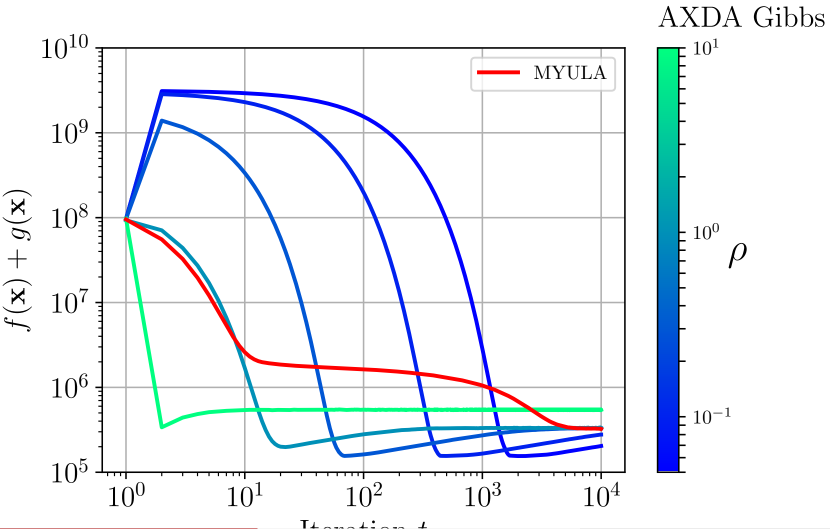
	SALSA	**FISTA**	**SGS**	**P-MYULA**
time (s)	1	10	470	3600
time (× var. split.)	1	10	**1**	7.7
nb. iterations	22	214	$\sim 10^4$	10^5
SNR (dB)	17.87	17.86	**18.36**	17.97

$$\text{Rk : } \rho^2 = 9$$

Short auto-correlation of the Markov chain



ρ = comput. time compromise/quality



Motivation for *augmentation*:

better mixing properties of the Markov chain

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

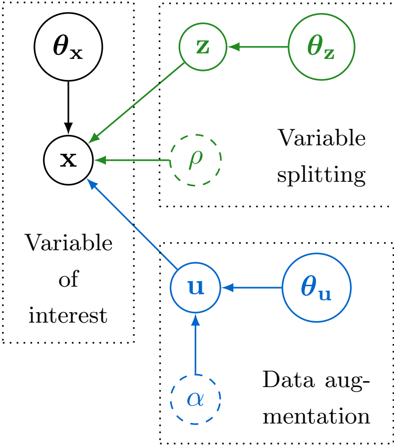
Assumption 2

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

$$\int_{\mathbb{R}^d} \exp\left[-\phi_1(\mathbf{x}, \mathbf{z}-\mathbf{u}; \rho) - \phi_2(\mathbf{u}; \alpha)\right] \mathrm{d}\mathbf{u} \propto \exp\left[-\phi_1(\mathbf{x}, \mathbf{z}; \eta(\rho, \alpha))\right]. \tag{1}$$

Splitted & Augmented Gibbs sampling (SPA)

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



The conditional split-augmented distributions are:

$$p(\mathbf{x}|\mathbf{z},\mathbf{u};\rho) \propto \exp\left[-f(\mathbf{x}) - \phi_1(\mathbf{x},\mathbf{z}-\mathbf{u};\rho)\right]$$

$$p(\mathbf{z}|\mathbf{x},\mathbf{u};\rho) \propto \exp\left[-g(\mathbf{z}) - \phi_1(\mathbf{x},\mathbf{z}-\mathbf{u};\rho)\right]$$

$$p(\mathbf{u}|\mathbf{x},\mathbf{z};\rho,\alpha) \propto \exp\left[-\phi_2(\mathbf{u};\alpha)\right] \times \exp\left[-\phi_1(\mathbf{x},\mathbf{z}-\mathbf{u};\rho)\right].$$

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:

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

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

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

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*

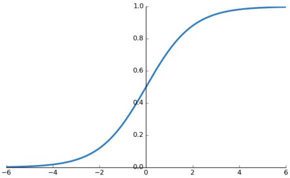
- Machine learning: logistic regression,...

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

Distributed sampling and data privacy

Regularized logistic regression

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



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

where

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

Issues:

- the full data set is distributed over b nodes, $b \in [\![1, n]\!]$
- data privacy.

Distributed sampling and data privacy

Applying AXDA b times

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

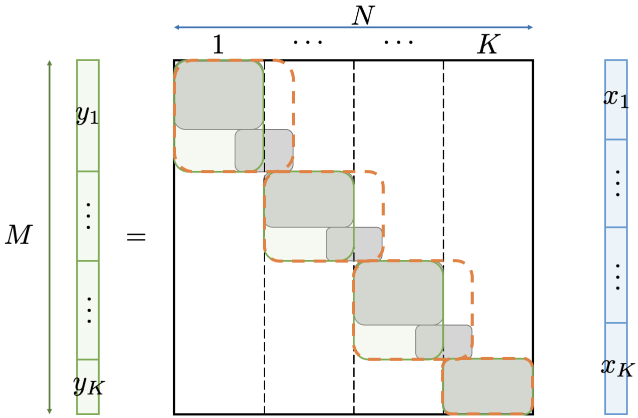
Benefits of AXDA:

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

Distributed sampling: much faster computations?

Particular case : **localized** observation operators

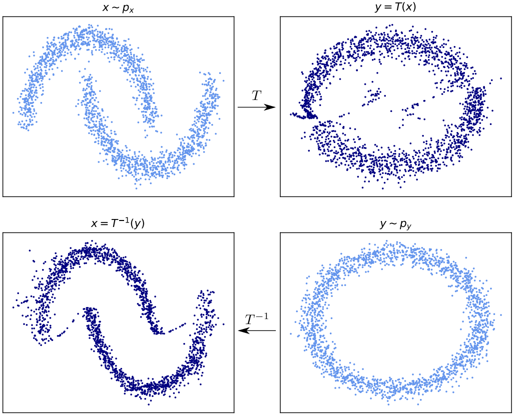
Collab. **Pierre-Antoine Thouvenin** & **Audrey Repetti**
(Edinburgh)



Beyond sampling: Toward generative models

(Deep) **Learning changes of variables**

Collab. **Florentin Cœurdoux** & **Nicolas Dobigeon** (IRIT)



Conclusion

- **SP & 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 (simulation, optimization, variational approx.)
- **AXDA**: **unifying** statistical framework
  - asymptotically exact: control parameter ρ
  - **non-asymptotic theoretical guarantees**

Sampling Gaussian variables in high dimensions

[Vono, C., Dobigeon, SIAM Review 2022]

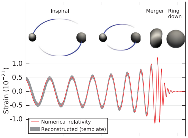
Prospects

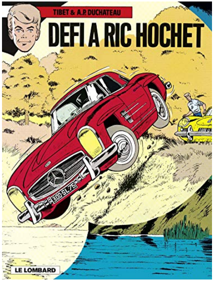
- **Distributed sampling: fast and scalable**
  - localized operators
  - distributed computing: coding
  - **confidence intervals**
- **Generative models** for inference
  - **learning** sampling networks
  - **evaluating** posterior distributions
- **AXDA**: **unifying** statistical framework
  - ELSA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

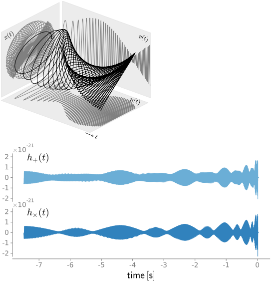
WP 2. Inverse problems and machine learning for polarized signals reconstruction

Task 2.1. Bayesian models: exploiting the geometry of the polarization information

Task 2.2. Machine learning for efficient Bayesian inference







Interested in AXDA for your statistical problems?

Theory and methods

- M. Vono et al. (2019), "Asymptotically exact data augmentation: models, properties and algorithms". Technical report. https://arxiv.org/abs/1902.05754/
- M. Vono et al. (2019), "Split-and-augmented Gibbs sampler - Application to large-scale inference problems," *IEEE Transactions on Signal Processing*.
- L. J. Rendell et al. (2018), "Global consensus Monte Carlo". Technical report. https://arxiv.org/abs/1807.09288/

Applications

- M. Vono et al. (2019), "Bayesian image restoration under Poisson noise and log-concave prior," in *Proc. ICASSP*.
- M. Vono et al. (2018), "Sparse Bayesian binary logistic regression using the split-and-augmented Gibbs sampler," in *Proc. MLSP*.

Code

- https://github.com/mvono

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.

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

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

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

Pereyra, M. (2016), "Proximal Markov chain Monte Carlo algorithms," *Statistics and Computing*, 26, 745–760.

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. (2018), "Global consensus Monte Carlo," [online]. Technical report. Available at `https://arxiv.org/abs/1807.09288/`.

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

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

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