HAL
open science

Preconditioned P-ULA for Joint Deconvolution-Segmentation of Ultrasound Images - Extended Version

Marie-Caroline Corbineau, Denis Kouamé, Emilie Chouzenoux, Jean-Yves Tourneret, Jean-Christophe Pesquet

To cite this version:

Marie-Caroline Corbineau, Denis Kouamé, Emilie Chouzenoux, Jean-Yves Tourneret, Jean-Christophe Pesquet. Preconditioned P-ULA for Joint Deconvolution-Segmentation of Ultrasound Images - Extended Version. IEEE Signal Processing Letters, 2019, 26 (10), pp.1456–1460. hal-02073283v4

HAL Id: hal-02073283

https://hal.science/hal-02073283v4

Submitted on 21 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Preconditioned P-ULA for Joint Deconvolution-Segmentation of Ultrasound Images – Extended Version

Marie-Caroline Corbineau, Denis Kouamé, Emilie Chouzenoux, Jean-Yves Tourneret, Jean-Christophe Pesquet

Abstract

Joint deconvolution and segmentation of ultrasound images is a challenging problem in medical imaging. By adopting a hierarchical Bayesian model, we propose an accelerated Markov chain Monte Carlo scheme where the tissue reflectivity function is sampled thanks to a recently introduced proximal unadjusted Langevin algorithm. This new approach is combined with a forward-backward step and a preconditioning strategy to accelerate the convergence, and with a method based on the majorization-minimization principle to solve the inner nonconvex minimization problems. As demonstrated in numerical experiments conducted on both simulated and *in vivo* ultrasound images, the proposed method provides high-quality restoration and segmentation results and is up to six times faster than an existing Hamiltonian Monte Carlo method.

Index Terms

Ultrasound, Markov chain Monte Carlo method, proximity operator, deconvolution, segmentation.

I. Introduction

In medical ultrasound (US) imaging, useful information can be drawn from the statistics of the tissue reflectivity function (TRF) to perform segmentation [1], tissue characterization [2], or classification [3]. Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ be the vectorized TRF and radio-frequency (RF) image, respectively. The following simplified model is used [4], [5]

$$y = Hx + \omega, \tag{1}$$

where $H \in \mathbb{R}^{n\times n}$ is a linear operator that models the convolution with the point spread function (PSF) of the probe, and $\omega \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_n)$, with $\mathcal{N}$ the normal distribution, and $\mathbb{I}_n$ the identity matrix in $\mathbb{R}^{n\times n}$. This paper assumes that the PSF is known, while $\sigma^2 > 0$ is an unknown parameter to be estimated. The TRF is comprised of K different tissues, which are identified by a hidden label field $z = (z_i)_{1\leq i\leq n} \in \{1, \dots, K\}^n$. For every $k \in \{1, \dots, K\}$, the kth region is modeled by a generalized Gaussian distribution ($\mathcal{GGD}$) [3], [6], which is parametrized by a shape parameter $\alpha_k \in [0, 3]$, related to the scatterer concentration, and a scale parameter $\beta_k > 0$, linked to the signal energy. Given y and H, the aim is to estimate a deblurred image x [7], [8], as well as σ^2, $\alpha = (\alpha_k)_{1\leq k\leq K}$, $\beta = (\beta_k)_{1\leq k\leq K}$, and the label field z. Due to the interdependence of these unknowns, it is beneficial to perform the deconvolution and segmentation tasks in a joint manner [9], [10]. This is achieved in [6] by considering a hierarchical Bayesian model, which is used within a Markov chain Monte Carlo (MCMC) method [11] to sample x, σ^2, α, β, and z according to the full conditional distributions. Despite promising results in image restoration and segmentation, the method in [6] is of significant computational complexity, in particular due to the adjusted Hamiltonian Monte Carlo (HMC) method [12], [13] used to sample the TRF. Recently, efficient and reliable stochastic sampling strategies have been devised [14], [15], [16] using the proximity operator [17], which is known as a useful tool for large-scale nonsmooth optimization [18]. In this work, we investigate an MCMC algorithm to perform the joint deconvolution and segmentation of US images, where the TRF is sampled with a scheme inspired from the proximal unadjusted Langevin algorithm (P-ULA) [15]. The latter generates samples according to an approximation of the target distribution without acceptance test, while being geometrically ergodic whereas classical unadjusted Langevin algorithms may have convergence issues.

A. Main contributions

Our contributions include *i*) the proposition of an original accelerated preconditioned version of P-ULA (PP-ULA), which relies on the use of a variable metric forward-backward strategy [19], [20], *ii*) an efficient solver based on the majorization-minimization (MM) principle to tackle the involved nonconvex priors, and *iii*) a new hybrid Gibbs sampler yielding a substantial reduction of the computational time needed to perform joint high-quality deconvolution and segmentation of both simulated and *in vivo* US images.

This article is organized as follows: Section II describes the investigated Bayesian model and sampling strategy. Section III focuses on the proposed TRF sampling method. Numerical experiments are finally presented in Section IV.

M.-C. Corbineau, E. Chouzenoux and J.-C. Pesquet are with the CVN, CentraleSupélec, INRIA Saclay, University Paris-Saclay, Gif-Sur-Yvette, France (e-mail: first-name.last-name@centralesupelec.fr)

D. Kouamé is with the IRIT, CNRS UMR 5505, University of Toulouse, Toulouse, France (e-mail: kouame@irit.fr).

J.-Y. Tourneret is with the IRIT, ENSEEIHT, TéSA, University of Toulouse, Toulouse, France (e-mail: Jean-Yves.Tourneret@enseeiht.fr).

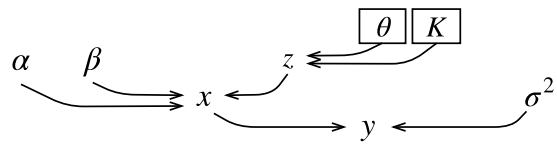


Fig. 1. Hierarchical Bayesian model. Parameters in boxes are fixed in advance.

II. Bayesian Model

A. Priors

Fig. 1 illustrates the hierarchical model used to perform a joint deconvolution-segmentation of ultrasound images. The following likelihood function is derived from (1)

$$p(y|x,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|y-Hx\|^2}{2\sigma^2}\right). \tag{2}$$

The TRF is a mixture of $\mathcal{GGD}$s which, under the assumption that the pixel values are independent given z, leads to

$$p(x|\alpha,\beta,z) = \prod_{i=1}^{n} \frac{1}{2\beta_{z_i}^{1/\alpha_{z_i}}\Gamma(1+1/\alpha_{z_i})} \exp\left(-\frac{|x_i|^{\alpha_{z_i}}}{\beta_{z_i}}\right). \tag{3}$$

Uninformative Jeffreys priors are assigned to the noise variance and scale parameters, while the shape parameters are assumed to be uniformly distributed between 0 and 3. The labels z are modeled by a Potts Markov random field with prior

$$p(z) = \frac{1}{C(\theta)} \exp\left(\sum\nolimits_{i=1}^{n} \sum\nolimits_{j\in\mathcal{V}(i)} \theta\delta(z_i - z_j)\right), \tag{4}$$

with δ the Kronecker function, $C(\theta) > 0$ a normalizing constant, $\theta > 0$ a granularity coefficient, and $\mathcal{V}(i)$ the set of four closest neighbours of the ith pixel.

B. Conditional distributions

The different variables are sampled according to their conditional distributions, which are provided in this section. The conditional distribution of the noise variance is derived from the Bayes theorem as follows

$$p(\sigma^2|y,x) \propto \mathcal{IG}\left(\frac{n}{2}, \frac{\|y-Hx\|^2}{2}\right), \tag{5}$$

where $\mathcal{IG}$ denotes the inverse gamma distribution. Assuming that the different regions have independent shape and scale parameters for every $k \in \{1,\ldots,K\}$, we obtain

$$p(\alpha_k|x,\beta,z) \propto \prod_{i\in\mathcal{I}_k} \frac{\mathbf{1}_{[0,3]}(\alpha_k)}{2\beta_k^{1/\alpha_k}\Gamma\left(1+1/\alpha_k\right)} \exp\left(-\frac{|x_i|^{\alpha_k}}{\beta_k}\right), \tag{6}$$

$$p(\beta_k|x,\alpha,z) \propto \mathcal{IG}\left(\frac{n_k}{\alpha_k}, \sum\nolimits_{i\in\mathcal{I}_k} |x_i|^{\alpha_k}\right), \tag{7}$$

with $\mathcal{I}_k = \{i \in \{1,\ldots,n\} \mid z_i = k\}$, n_k the number of elements in $\mathcal{I}_k$, and $\mathbf{1}_{[0,3]}$ the characteristic function of $[0,3]$. Samples for α_k are drawn from (6) by using a Metropolis-Hastings (MH) random walk. For every pixel $i \in \{1,\ldots,n\}$ and every region $k \in \{1,\ldots,K\}$, the Bayes rule applied to the segmentation labels leads to

$$p(z_i = k|x,\alpha,\beta,z_{\mathcal{V}(i)}) \propto \frac{\exp\left(\sum_{j\in\mathcal{V}(i)} \theta\delta(z_j - k) - \frac{|x_i|^{\alpha_k}}{\beta_k}\right)}{2\beta_k^{1/\alpha_k}\Gamma(1+1/\alpha_k)} \tag{8}$$

where $z_{\mathcal{V}(i)}$ denotes the label values in the neighborhood of i. As a consequence, the label z_i is drawn from $\{1,\ldots,K\}$ using the above probabilities (suitably normalized).

III. Preconditioned P-ULA

A. Notation

Let $\mathcal{I}_{<1} = \{i \in \{1,\dots,n\} \mid \alpha_{z_i} < 1\}$ and $\mathcal{I}_{\geq 1} = \{1,\dots,n\} \setminus \mathcal{I}_{<1}$. Let $\mathcal{S}_n$ denote the set of symmetric positive definite matrices in $\mathbb{R}^{n\times n}$, and let $|||\cdot|||$ denote the spectral norm. For every $Q \in \mathcal{S}_n$, let $\|\cdot\|_Q = \langle \cdot, Q\cdot\rangle^{1/2}$. For every function $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$, the proximity operator of f at $x \in \mathbb{R}^n$ with respect to the norm induced by $Q^{-1} \in \mathcal{S}_n$ is defined as follows [17],

$$\mathrm{prox}_f^Q(x) \in \mathrm{Argmin}_{u\in\mathbb{R}^n} \frac{1}{2}\|x-u\|^2_{Q^{-1}} + f(u). \tag{9}$$

If Q is not specified, then $Q = \mathbb{I}_n$. If prox_f is simple to compute, then the solution to (9) for an arbitrary $Q \in \mathcal{S}_n$ can be obtained by using the dual forward-backward (DFB) algorithm [21], summarized in Algorithm 1. If f is proper, lower semicontinuous, and convex, then the sequence $(u^{(p)})_{p\in\mathbb{N}}$ generated by Algorithm 1 converges to $\mathrm{prox}_f^Q(x)$.

Algorithm 1: DFB algorithm to compute $\mathrm{prox}_f^Q(x)$

Initialize dual variable $w^{(1)} \in \mathbb{R}^N$;
Set $\rho = |||Q|||^{-1}$, $\epsilon \in]0, \min\{1,\rho\}[$, $\eta \in [\epsilon, 2\rho-\epsilon]$;
for $p = 1, \dots$ **do**
  $u^{(p)} = x - Qw^{(p)}$;
  $w^{(p+1)} = w^{(p)} + \eta u^{(p)} - \eta \mathrm{prox}_{\eta^{-1}f}(\eta^{-1}w^{(p)} + u^{(p)})$
end

B. Sampling the TRF

The conditional distribution of the TRF is

$$\pi(x) = p(x|y,\sigma^2,\alpha,\beta,z) \propto \exp\left(-\frac{\|y-Hx\|^2}{2\sigma^2} - g(x)\right), \tag{10}$$

where $(\forall x \in \mathbb{R}^n)\ g(x) = \sum_{i=1}^n \beta_{z_i}^{-1}|x_i|^{\alpha_{z_i}}$. Let $\gamma > 0$ and let $Q \in \mathcal{S}_n$ be a preconditioning matrix used to accelerate the sampler [22]. Following [15], $\pi(x)$ is approximated by

$$\pi_\gamma(x) \propto \sup_{u\in\mathbb{R}^n} \pi(u) \exp\left(-\frac{\|u-x\|^2_{Q^{-1}}}{2\gamma}\right). \tag{11}$$

As shown in Appendix, the Euler discretization of the Langevin diffusion equation [23] applied to π_γ with stepsize 2γ and preconditioning matrix Q leads to

$$x^{(t+1)} = \mathrm{prox}_{\gamma g}^Q(\tilde{x}^t) + \sqrt{2\gamma}Q^{\frac{1}{2}}\omega^{(t+1)}, \tag{12}$$

where $\omega^{(t+1)} \sim \mathcal{N}(0,\mathbb{I}_n)$ and

$$\tilde{x}^{(t)} = x^{(t)} - \frac{\gamma}{\sigma^2}QH^\top(Hx^{(t)} - y). \tag{13}$$

Since the proposed sampling strategy is unadjusted, (12) is not followed by an acceptance test. The bias with respect to π increases with γ, as the speed of convergence of the algorithm. A compromise must be found when setting γ.

When $\mathcal{I}_{<1}$ is not empty, we use the MM principle [24] to replace the nonconvex minimization problem involved in the computation of $\mathrm{prox}_{\gamma g}^Q$ with a sequence of convex surrogate problems. Let $\mathcal{J} \subset \mathcal{I}_{<1}$. We define $h_\mathcal{J}$ at every $(u,v) \in \mathbb{R}^n \times \mathbb{R}^n_{+*}$ by

$$h_\mathcal{J}(u,v) = \sum_{i\in\mathcal{I}_{\geq 1}} \frac{|u_i|^{\alpha_{z_i}}}{\beta_{z_i}} + \sum_{j\in\mathcal{J}} \frac{(1-\alpha_{z_j})v_j^{\alpha_{z_j}} + \alpha_{z_j}v_j^{\alpha_{z_j}-1}|u_j|}{\beta_{z_j}}.$$

From concavity, we deduce that, for every $v \in \mathbb{R}^n_{+*}$ and $u \in \mathbb{R}^n$ such that $\mathcal{J} \subset \{i \in \mathcal{I}_{<1} \mid |u_i| > 0\}$, the following majoration property holds

$$h_\mathcal{J}(u,v) \geq \sum_{i\in\mathcal{I}_{\geq 1}\cup\mathcal{J}} \frac{|u_i|^{\alpha_{z_i}}}{\beta_{z_i}} = h_\mathcal{J}(u,(|u_i|)_{1\leq i\leq n}).$$

Since $h_\mathcal{J}(\cdot,v)$ is convex and separable, its proximity operator in the Euclidean metric is straightforward to compute. More precisely, for every $i \in \mathcal{I}_{\geq 1}$, $\eta > 0$ and $s \in \mathbb{R}$, $\mathrm{prox}_{\eta^{-1}|\cdot|^{\alpha_{z_i}}}(s)$ has either a closed form [25] or can be found using a bisection search in $[0,|s|]$. Algorithm 1 can then be called, in order to compute the proximity operator of $h_\mathcal{J}(\cdot,v)$ in any metric $Q \in \mathcal{S}_n$. This leads to Algorithm 2 which generates a sequence $(u^{(q)})_{q\in\mathbb{N}}$ estimating $\mathrm{prox}_{\gamma g}^Q(\tilde{x}^{(t)})$.

The resulting Gibbs sampler is summarized in Algorithm 3.

Algorithm 2: MM principle to compute $\mathrm{prox}_{\gamma g}^{Q}$.

Initialize $u^{(1)} \in \mathbb{R}^n$;
for $q = 1, \ldots$ **do**
  $\mathcal{J}^{(q)} = \{i \in \mathcal{I}_{<1} \mid |u_i^{(q)}| > 0\}$;
  $v^{(q)} = (|u_i^{(q)}|)_{1 \leq i \leq n}$;
  $u^{(q+1)} = \mathrm{prox}^{Q}_{\gamma h_{\mathcal{J}^{(q)}}(\cdot, v^{(q)})}(\tilde{x}^{(t)})$ (using Alg. 1)
end

Algorithm 3: Hybrid Gibbs sampler

1 Sample the noise variance σ^2 according to (5);
2 Sample the shape parameter α using MH with (6);
3 Sample the scale parameter β according to (7);
4 Sample the hidden label field z using (8) ;
5 Sample the TRF x using PP-ULA (12)-(13).

IV. Numerical experiments

A. Experimental settings

Six experiments are presented. Simu1 and Simu2 refer to simulated images with two and three regions, respectively. Kidney denotes the tissue-mimicking phantom produced from 10^6 scatterers uniformly distributed over a digital image of human kidney tissue provided with the Field II ultrasound simulator [26]. The amplitude of each scatterer is produced using a zero-mean Gaussian distribution whose variance is linked to the amplitude of the point on the digital image. The PSF for the aforementioned simulations is obtained with Field II and corresponds to a 3.5 MHz linear probe. We also perform tests on three real ultrasound images. Thyroid denotes a real RF image of thyroidal flux obtained *in vivo* with a 7.8 MHz probe. The unknown PSF is identified using the RF image of a wire cross-section which was acquired with the same probe. Since the diameter of the wire is of the order of a few μm, its cross-section can almost be viewed as a point. Thus, its RF image provides a good approximation of the PSF. Finally, Bladder and KidneyReal refer to the RF images of a mouse bladder and mouse kidney, respectively. Both images were obtained *in vivo* with a 20 MHz probe. The PSF for these two real images is estimated using the same method as for Thyroid. The number of regions K is set to 2 for Simu1 and KidneyReal, and it is set to 3 for Simu2, Kidney, Thyroid and Bladder. The test settings and images can be found in Table I and Figs. 4 and 6 (first column), respectively.

The TRF is initialized using a pre-deconvolved image obtained with a Wiener filter, while the segmentation is initialized by applying a 7×7 median filter and the Otsu method [27] to the B-mode of the initial TRF. Shape and scale parameters are randomly selected in $[0.5, 1.5]$, and $[1, 200]$, respectively. The granularity parameter θ for the Potts model (4) is adjusted to ensure that the percentage of isolated points in the segmentation, obtained with a 3×3 median filter, is close to 0.05, 0.1, 0.8, 0.08, 0.08 and 0.08 for Simu1, Simu2, Kidney, Thyroid, Bladder and KidneyReal, respectively.

B. Comparisons and evaluation metrics

All computational times are given for simulations run on Matlab 2018b on an Intel Xeon CPU E5-1650 3.20GHz. The code for the proposed method is available online[1]. In addition to comparing Algorithm 3 with HMC [6], the quality of the deconvolution is compared with the one obtained with a Wiener filter, where the noise level has been estimated as in [28], and with the solution to the Lasso problem, where the regularization weight is set *i*) manually when the ground-truth is not

[1] https://github.com/mccorbineau/PP-ULA

		Simu1	Simu2	Kidney	Thyroid	Bladder	KidneyReal
	Size	256×256	256×256	294×354	870×140	370×256	350×200
	Data type	Simulated	Simulated	Tissue-mimicking	Real *in vivo*	Real *in vivo*	Real *in vivo*
Ground-truth	TRF	✓	✓	✓	-	-	-
	$\mathcal{GGD}$ parameters	✓	✓	-	-	-	-
	Segmentation	✓	✓	-	-	-	-

TABLE I
Test settings: size of test images, data type, and availability of the ground-truth.

available, or *ii*) using a golden-section search to maximize the peak signal-to-noise ratio (PSNR) defined as (with x^{tr} the true TRF and x^{es} the estimated one)

$$\mathrm{PSNR} = 10\log_{10}(n\ \max_i(x_i^{\mathrm{tr}}, x_i^{\mathrm{es}})^2/\|x^{\mathrm{tr}} - x^{\mathrm{es}}\|^2). \tag{14}$$

We also compare our results with the segmentation given by Otsu's method [27] applied to the Wiener-deconvolved image, and with the SLaT method [29] applied to the Lasso-deconvolved image. PP-ULA is used with $\gamma = 0.09$ and Q an approximation of the inverse of the Hessian of the differentiable term in (10) [30], $Q = \sigma^2(H^\top H + \lambda \mathbb{I}_n)^{-1}$, with $\lambda = 0.1$ so that Q is well-defined. We have also computed the structural similarity measure (SSIM) [31] of the restored TRF and the contrast-to-noise ratio (CNR) [32] between two windows from different regions of the B-mode TRF images. The segmentation is evaluated according to the percentage of correctly predicted labels, or overall accuracy (OA). The minimum mean square error (MMSE) estimators of all parameters in HMC and PP-ULA are computed after the burn-in regime. Moreover, to evaluate the mixing property of the Markov chain after convergence, we compute the mean square jump (MSJ) per second, which is the ratio of the MSJ to the time per iteration. The MSJ is obtained using T samples of the TRF $(x^{t_0+1}, \ldots, x^{t_0+T})$ generated after the burn-in period, i.e. $\mathrm{MSJ} = \left(\frac{1}{T-1}\sum_{t=1}^{T-1}\|x^{(t_0+t)} - x^{(t_0+t+1)}\|^2\right)^{1/2}$.

C. Results on simulated data

The convergence speed of Algorithm 3 is empirically observed for Simu1 and Simu2, as illustrated in Fig. 2, where we also display the results of the non-preconditioned P-ULA, for which $Q = \mathbb{I}_n$ and $\gamma = 1.99\sigma^2/|||H|||^2$. Comparing P-ULA and PP-ULA on these simulated data allows us to study the effect of adding a preconditioner in the proposed sampling scheme. As reported in Table II, P-ULA needs more iterations and more time to converge than PP-ULA: the proposed method is 12.2 and 4.8 times faster than P-ULA on Simu1 and Simu2, respectively. In addition, from Table III and Fig. 3, we deduce that P-ULA is more biased than PP-ULA, which samples correctly the target distributions. Finally, as one can see in Fig. 2 and Table IV, P-ULA leads to lower PSNR, SSIM and OA values than PP-ULA. These results clearly emphasize the benefits of preconditioning in this example.

		Iterations		Time		Mixing property
		Burn-in	Total	Duration	PP-ULA speed gain	MSJ (per s)
Simu1	P-ULA	70000	140000	2 h 27 min	12.2	665
	HMC	4000	8000	1 h 08 min	5.7	173
	PP-ULA	4000	8000	12 min	1	970
Simu2	P-ULA	70000	140000	3 h 06 min	4.8	590
	HMC	10000	20000	4 h 14 min	6.6	22
	PP-ULA	10000	20000	39 min	1	793

TABLE II
NUMBER OF ITERATIONS, COMPUTATIONAL TIME AND MSJ PER S FOR EXPERIMENTS SIMU1 AND SIMU2.

	Simu1					Simu2						
	σ^2	α_1	β_1	α_2	β_2	σ^2	α_1	β_1	α_2	β_2	α_3	β_3
True	0.013	1.5	1.0	0.60	1.0	33	1.5	100	1.0	50	0.50	4.0
P-ULA	0.041	2.0	0.5	0.59	1.0	122	2.0	330	2.0	3186	0.48	3.4
HMC	0.013	1.8	1.2	0.61	1.0	34	1.4	66	1.1	111	0.54	5.2
PP-ULA	0.013	1.4	0.9	0.62	1.1	35	2.3	2676	1.2	122	0.55	5.8

TABLE III
MMSE ESTIMATES OF THE NOISE VARIANCE AND $\mathcal{GGD}$ PARAMETERS.

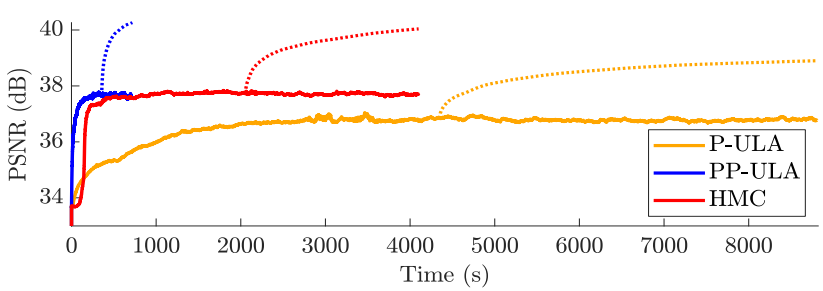


Fig. 2. PSNR along time for Simu1. Dotted lines indicate the PSNR of the MMSE estimator of the TRF after the burn-in regime.

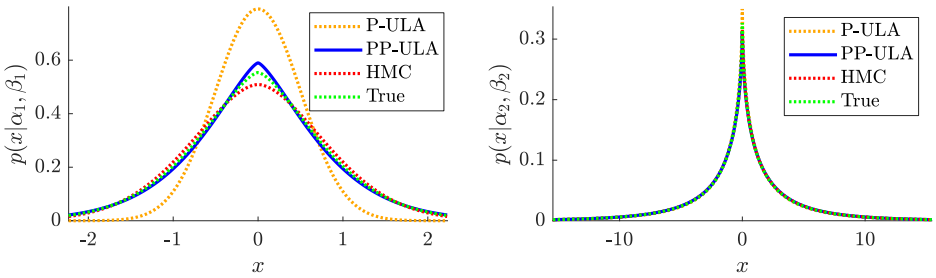


Fig. 3. Simu1, $\mathcal{GGD}$ distributions (3) of regions 1 (left) and 2 (right).

	Simu1				Simu2			
	PSNR	SSIM	CNR	OA	PSNR	SSIM	CNR	OA
Wiener - Otsu	37.1	0.57	1.26	99.5	35.4	0.63	0.97	96.0
Lasso - SLaT [29]	39.2	0.60	1.15	99.6	37.8	0.70	0.99	98.3
P-ULA	38.9	0.45	**1.82**	98.7	37.1	0.57	1.59	94.9
HMC	40.0	**0.62**	1.47	**99.7**	36.4	0.64	1.59	98.5
PP-ULA	**40.3**	**0.62**	1.51	**99.7**	**38.6**	**0.71**	**1.64**	**98.7**

TABLE IV
PSNR, SSIM, CNR AND SEGMENTATION OVERALL ACCURACY (OA) FOR SIMULATED DATA.

From Table II, PP-ULA is 5.7 and 6.6 times faster than HMC on Simu1 and Simu2 and has better mixing properties, as shown by the MSJ per second. Visual results from Fig. 4 and CNR values in Table IV show that the contrast obtained with PP-ULA is better than with competitors on Simu2, and is second best after P-ULA on Simu1. However, it should be noted that the PSNR and SSIM obtained on Simu1 with P-ULA are much lower than with the other methods. In addition, the PSNR and SSIM values from Table IV obtained with PP-ULA are equivalent or higher than all competitors for these two experiments. Visual segmentation results are shown in Fig. 5, and OA values can be found in Table IV. For these simulated images, more pixels are correctly labeled with PP-ULA than with competitors.

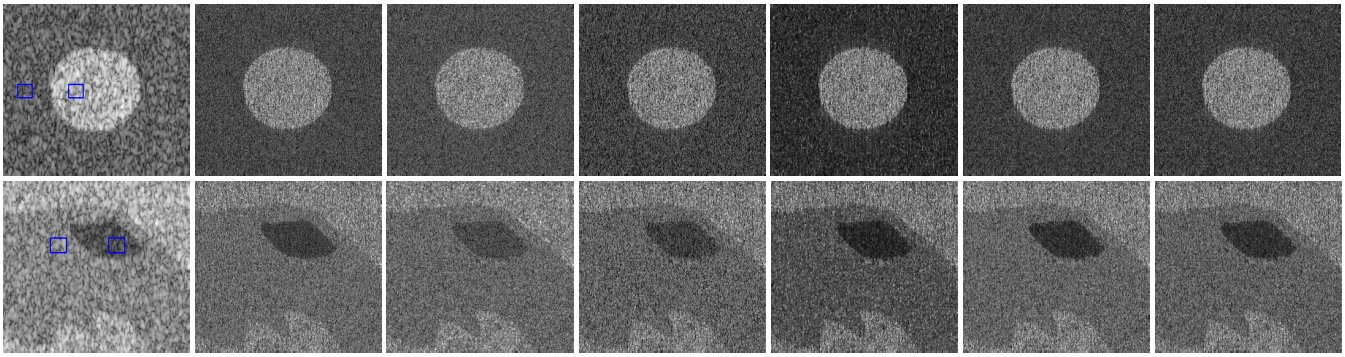


Fig. 4. B-mode visualization on simulated images. Top: Simu1. Bottom: Simu2. Left to right: RF image, TRF: ground-truth, Wiener, Lasso, P-ULA, HMC, PP-ULA. Blue boxes indicate regions used for the CNR.

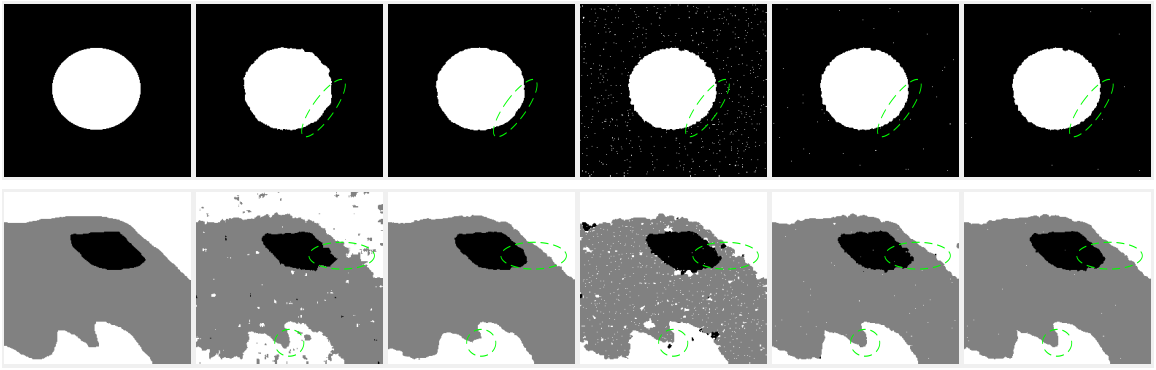


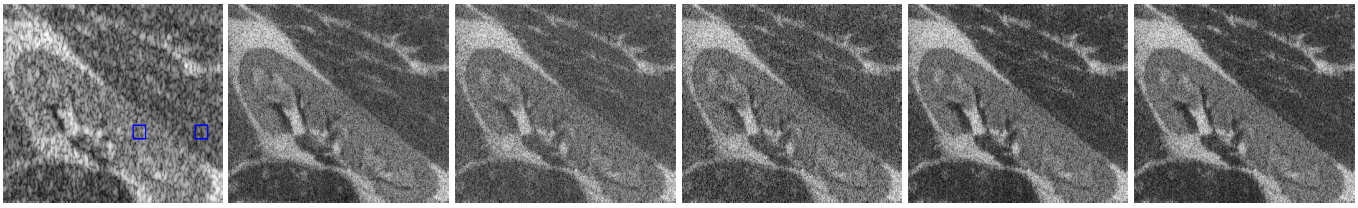
Fig. 5. Segmentation. Top: Simu1, bottom: Simu2. Left to right: ground-truth, Otsu, SLaT, P-ULA, HMC, PP-ULA. Main differences are circled in green.

		Iterations		Time		Mixing property
		Burn-in	Total	Duration	PP-ULA speed gain	MSJ (per s)
Kidney	HMC	7000	14000	4 h 23 min	6.3	167
	PP-ULA	7000	14000	42 min	1	657
Thyroid	HMC	3000	6000	2 h 09 min	3.7	175
	PP-ULA	3000	6000	35 min	1	950
Bladder	HMC	5000	10000	2 h 45 min	5.2	13
	PP-ULA	5000	10000	32 min	1	1396
KidneyReal	HMC	5000	10000	1 h 49 min	5.8	11
	PP-ULA	5000	10000	19 min	1	1361

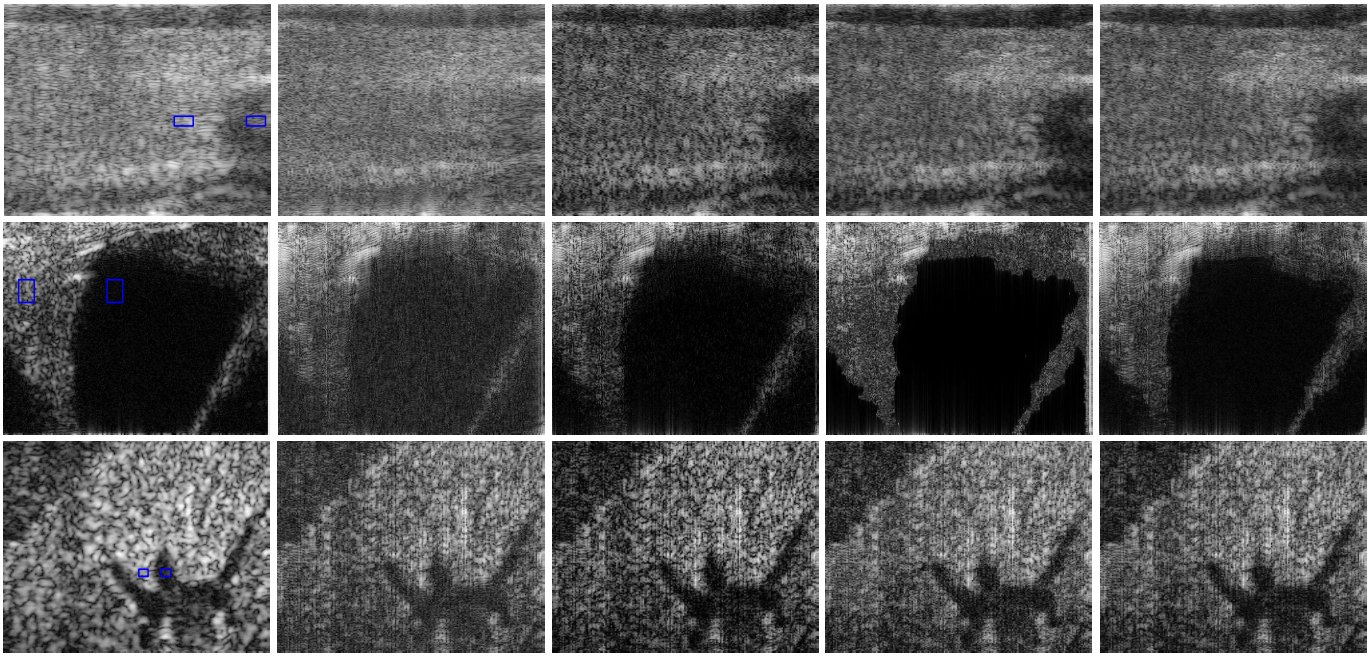
TABLE V
NUMBER OF ITERATIONS, COMPUTATIONAL TIME AND MSJ PER S FOR EXPERIMENTS ON THE TISSUE-MIMICKING PHANTOM AND ON REAL DATA.

D. Results on a tissue-mimicking phantom and on real data

The convergence of Algorithm 3 is also empirically observed for the experiments on the tissue-mimicking phantom and on real data, i.e. Kidney, Thyroid, Bladder and KidneyReal. As mentioned in Table V, the proposed method leads to a significant acceleration since it is between 3.7 and 6.3 times faster than HMC on these experiments. Visual results from Fig. 6 and CNR values in Table VI show that the contrast obtained with PP-ULA is better than with competitors on all these test images. In addition, the PSNR and SSIM values from Table VI obtained with PP-ULA on the Kidney experiment are equivalent or higher than all competitors. Although the ground-truth of the segmentation is not available for these experiments, one can see from the visual segmentation results shown in Fig. 7, that the segmentation based on the Potts model (PP-ULA and HMC) gives more homogeneous areas than Otsu, and recovers more details than SLaT.



(a)



(b)

Fig. 6. B-mode visualization. (a) Kidney; left to right: RF image, TRF: ground-truth, Wiener, Lasso, HMC, PP-ULA. (b) Real *in vivo* images; top to bottom: Thyroid, Bladder, KidneyReal; left to right: RF image, TRF: Wiener, Lasso, HMC, PP-ULA. Blue boxes indicate regions used for the CNR.

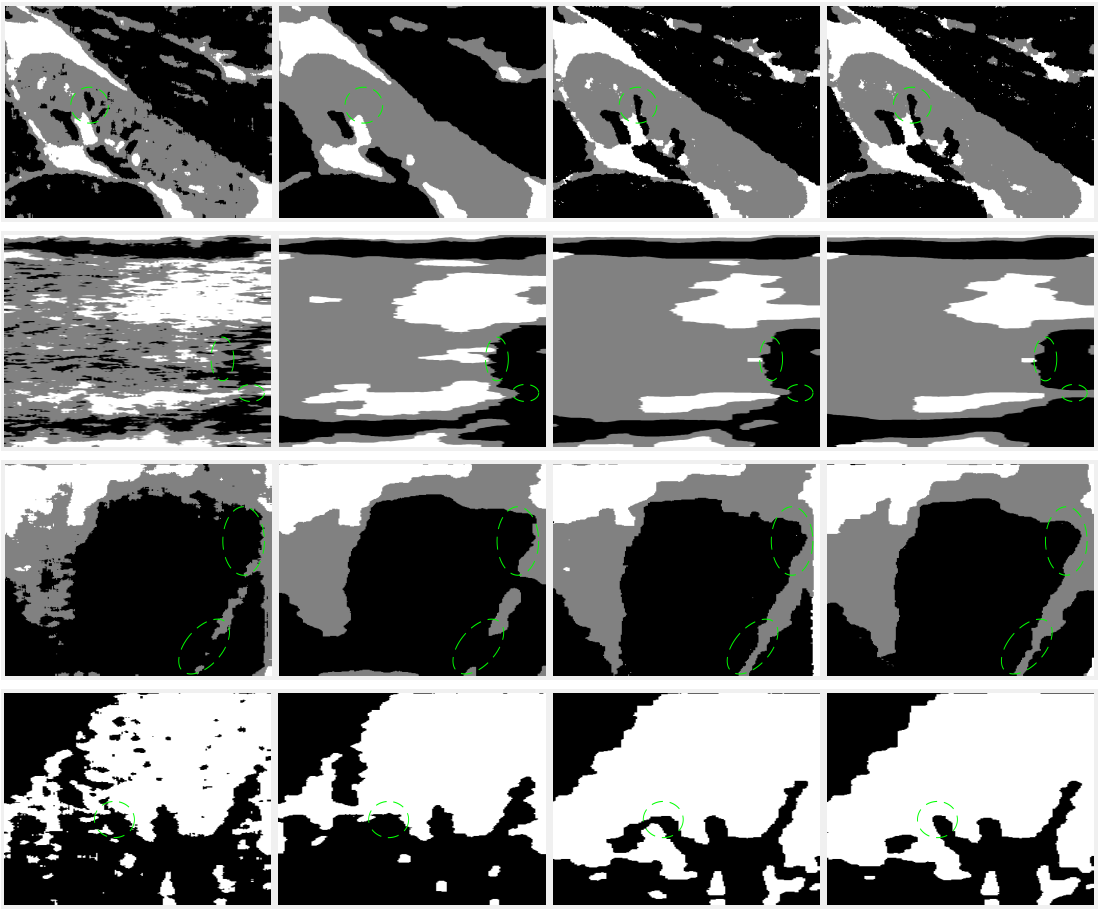


Fig. 7. Segmentation. Top to bottom: Kidney, Thyroid, Bladder, KidneyReal. Left to right: Otsu, SLaT, HMC, PP-ULA. Main differences are circled in green.

	Kidney			Thyroid	Bladder	KidneyReal
	PSNR	SSIM	CNR	CNR	CNR	CNR
Wiener	27.6	0.58	0.66	0.56	1.66	1.61
Lasso	28.5	0.59	0.67	0.99	1.76	1.76
HMC	**29.5**	**0.62**	1.10	1.52	2.23	1.88
PP-ULA	29.3	**0.62**	**1.14**	**1.56**	**2.48**	**1.93**

TABLE VI
PSNR, SSIM AND CNR RESULTS.

V. CONCLUSION

We investigated a new method based on a preconditioned proximal unadjusted Langevin algorithm for the joint restoration and segmentation of ultrasound images, which showed faster convergence than an existing Hamiltonian Monte Carlo algorithm. Hence, the proposed method has the potential to speed-up the approach proposed in [1] for the segmentation of ultrasound images. Another direction for future work is to extend this framework to a spatially variant, possibly unknown, PSF.

APPENDIX

In this section, after reminding results about the Langevin diffusion and its discretization using Euler's scheme, we provide details about the derivation of the proposed method (12) used to sample the TRF.

A. Discrete Langevin diffusion

An n-dimensional Langevin diffusion is a continuous time Markov process $(x(t))_{t\in[0,+\infty[}$ taking its values in $\mathbb{R}^n$, which is the solution to the following stochastic differential equation [23],

$$(\forall t \in [0,+\infty[) \quad \mathrm{d}x(t) = b(x(t))\mathrm{d}t + V(x(t))\mathrm{d}B(t), \tag{15}$$

where $(B(t))_{t\in[0,+\infty[}$ is a Brownian motion with values in $\mathbb{R}^n$, and for every $x \in \mathbb{R}^n$, $V(x) \in \mathbb{R}^{n\times n}$ is the volatility matrix and $b(x) = (b_i(x))_{1\leq i\leq n} \in \mathbb{R}^n$ is the drift term defined as

$$(\forall i \in \{1,\dots,n\}) \quad b_i(x) = \frac{1}{2}\sum_{j=1}^{n} A_{i,j}(x)\frac{\partial \log \pi(x)}{\partial x_j} + \det(A(x))^{\frac{1}{2}} \sum_{k=1}^{n} \frac{\partial}{\partial x_j}\left(A_{i,k}(x)\det(A(x))^{-\frac{1}{2}}\right), \tag{16}$$

where $A(x) = V(x)V(x)^\top = (A_{i,j}(x))_{1\leq i,j\leq n}$ is a symmetric positive definite matrix, $\det(A(x))$ denotes its determinant, and π is the density of the stationary distribution of the diffusion. Here, we take $(\forall x \in \mathbb{R}^n)\ \pi(x) = p(x|y,\sigma^2,\alpha,\beta,z)$ defined in (10). Euler's discretization scheme applied to (15) leads to the following target posterior distribution, which can be used to generate a Langevin Markov chain.

$$(\forall t \in \mathbb{N}) \quad x^{(t+1)} = x^{(t)} + 2\gamma b(x^{(t)}) + \sqrt{2\gamma}A^{\frac{1}{2}}(x^{(t)})\omega^{(t)}. \tag{17}$$

Hereabove, $\omega^{(t)} \sim \mathcal{N}(0,\mathbb{I}_n)$ and $\gamma > 0$ is the discretization stepsize that controls the length of the jumps, while the scale matrix $A(\cdot)$ drives their direction. Instead of taking $A(\cdot) = \mathbb{I}_n$ as in the standard Metropolis adjusted Langevin algorithm, we follow [22], [19] and use a preconditioning matrix A to accelerate the Langevin scheme, which leads to

$$x^{(t+1)} = x^{(t)} + \gamma A\nabla \log \pi(x) + \sqrt{2\gamma}A^{\frac{1}{2}}\omega^{(t+1)}. \tag{18}$$

B. Approximation of the target diffusion

For every $x \in \mathbb{R}^n$, let $f(x) = \|y - Hx\|^2/(2\sigma^2)$. From (10), the target distribution statisfies the following relation,

$$(\forall x \in \mathbb{R}^n) \quad \pi(x) = p(x|y,\sigma^2,\alpha,\beta,z) \propto \exp(-(f+g)(x)), \tag{19}$$

where $(\forall x \in \mathbb{R}^n)\ g(x) = \sum_{i=1}^{n} \beta_{z_i}^{-1}|x_i|^{\alpha_{z_i}}$. Let $\gamma > 0$ and $Q \in \mathcal{S}_n$. Following [15], we replace π by its Moreau approximation π_γ^Q defined in (11) and recalled below,

$$(\forall x \in \mathbb{R}^n) \quad \pi_\gamma^Q(x) = \sup_{u\in\mathbb{R}^n} \pi(u)\exp\left(-\frac{\|u-x\|_{Q^{-1}}^2}{2\gamma}\right). \tag{20}$$

Note that we dropped the normalization constant. Moreover, at the difference of [15], we introduce the preconditioning matrix Q for convergence acceleration purposes. When Q is not specified, the identity matrix is used, *i.e.* $Q = \mathbb{I}_n$. Hence, the approximated version of (18) reads

$$x^{(t+1)} = x^{(t)} + \gamma A\nabla \log \pi_\gamma^Q(x) + \sqrt{2\gamma}A^{\frac{1}{2}}\omega^{(t+1)}. \tag{21}$$

We can then deduce the following result when g is convex.

Proposition 1: For every $\gamma > 0$, $Q \in \mathcal{S}_n$ and $x \in \mathbb{R}^n$, if $(\forall k \in \{1,\dots,K\})\ \alpha_k \geq 1$, then we have

$$\nabla \log \pi_\gamma^Q(x) = Q^{-1}\frac{\mathrm{prox}_{\gamma(f+g)}^Q(x) - x}{\gamma}. \tag{22}$$

Proof. By definition of π_γ^Q, we have

$$(\forall x \in \mathbb{R}^n) \quad \log \pi_\gamma^Q(x) = -(f+g)\left(\mathrm{prox}_{\gamma(f+g)}^Q(x)\right) - \frac{1}{2\gamma}\left\|\mathrm{prox}_{\gamma(f+g)}^Q(x) - x\right\|_{Q^{-1}}^2. \tag{23}$$

Hence, applying [33, Lemma 2.5] in the metric induced by Q^{-1} directly leads to the result. □

From Proposition 1, (21) becomes

$$x^{(t+1)} = x^{(t)} + \gamma AQ^{-1}\frac{\mathrm{prox}_{\gamma(f+g)}^Q(x^{(t)}) - x^{(t)}}{\gamma} + \sqrt{2\gamma}A^{\frac{1}{2}}\omega^{(t+1)}. \tag{24}$$

It can be noted that, in Proposition 1, g is assumed to be convex, which is not necessarily satified in our case. However, for simplicity, we take the discrete scheme (24) even in the nonconvex case. Finally, we take $A = Q$, which leads to

$$x^{(t+1)} = \mathrm{prox}_{\gamma(f+g)}^Q(x^{(t)}) + \sqrt{2\gamma}Q^{\frac{1}{2}}\omega^{(t+1)}. \tag{25}$$

C. Forward-backward approximation

By definition, f is differentiable on $\mathbb{R}^n$ and its gradient $\nabla f = H^\top(H \cdot -y)/\sigma^2$ is Lipschitz-continuous on $\mathbb{R}^n$. It is worth noting that the computation of the proximity operator of the sum of two functions is generally intractable [34]. Hence, as suggested in [15], we use a first-order Taylor expansion to approximate the proximity operator of $f+g$ and introduce a forward step in PP-ULA iteration. Let o denotes Landau's notation.[2] Let $x \in \mathbb{R}^n$, using $(\forall u \in \mathbb{R}^n)\ f(u) = f(x) + (u-x)^\top \nabla f(x) + o(\|u-x\|)$, we have

$$(f+g)(u) + \frac{1}{2\gamma}\|u-x\|^2_{Q^{-1}} = f(x) + g(u) + \frac{1}{2\gamma}\|u-x\|^2_{Q^{-1}} + (u-x)^\top \nabla f(x) + o(\|u-x\|), \tag{26}$$

which can be re-written as

$$(f+g)(u) + \frac{1}{2\gamma}\|u-x\|^2_{Q^{-1}} = f(x) + g(u) + \frac{1}{2\gamma}\|u-x+\gamma Q\nabla f(x)\|^2_{Q^{-1}} - \frac{\gamma}{2}\|Q^{\frac{1}{2}}\nabla f(x)\|^2 + o(\|u-x\|). \tag{27}$$

Hence, the proximity operator of $f+g$ can be expressed as follows,

$$\begin{aligned} \mathrm{prox}^Q_{\gamma(f+g)}(x) &= \underset{u\in\mathbb{R}^n}{\mathrm{argmin}} \left((f+g)(u) + \frac{1}{2\gamma}\|u-x\|^2_{Q^{-1}} \right) && (28)\\ &= \underset{u\in\mathbb{R}^n}{\mathrm{argmin}} \left(g(u) + \frac{1}{2\gamma}\|u-x+\gamma Q\nabla f(x)\|^2_{Q^{-1}} + o(\|u-x\|) \right). && (29) \end{aligned}$$

In addition, we have

$$\mathrm{prox}^Q_{\gamma g}(x - \gamma Q\nabla f(x)) = \underset{u\in\mathbb{R}^n}{\mathrm{argmin}} \left(g(u) + \frac{1}{2\gamma}\|u-x+\gamma Q\nabla f(x)\|^2_{Q^{-1}} \right). \tag{30}$$

Therefore, when γ is small, $\mathrm{prox}^Q_{\gamma g}(x - \gamma Q\nabla f(x))$ is a good approximation of $\mathrm{prox}^Q_{\gamma(f+g)}(x)$. Plugging this preconditioned forward-backward scheme [18] in (25) leads to the proposed sampling method

$$x^{(t+1)} = \mathrm{prox}^Q_{\gamma g}(x^{(t)} - \gamma Q\nabla f(x^{(t)})) + \sqrt{2\gamma}Q^{\frac{1}{2}}\omega^{(t+1)}. \tag{31}$$

References

[1] M. A. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret, "Segmentation of skin lesions in 2D and 3D ultrasound images using a spatially coherent generalized Rayleigh mixture model," *IEEE Transactions on Medical Imaging*, vol. 31, no. 8, pp. 1509–1520, 2012.

[2] O. Bernard, J. D'hooge, and D. Friboulet, "Statistics of the radio-frequency signal based on K distribution with application to echocardiography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 53, no. 9, pp. 1689–1694, 2006.

[3] M. Alessandrini, S. Maggio, J. Porée, L. De Marchi, N. Speciale, E. Franceschini, O. Bernard, and O. Basset, "A restoration framework for ultrasonic tissue characterization," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 58, no. 11, 2011.

[4] J. A. Jensen, J. Mathorne, T. Gravesen, and B. Stage, "Deconvolution of in-vivo ultrasound B-mode images," *Ultrasonic Imaging*, vol. 15, no. 2, pp. 122–133, 1993.

[5] J. Ng, R. Prager, N. Kingsbury, G. Treece, and A. Gee, "Modeling ultrasound imaging as a linear, shift-variant system," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 53, no. 3, pp. 549–563, 2006.

[6] N. Zhao, A. Basarab, D. Kouamé, and J.-Y. Tourneret, "Joint segmentation and deconvolution of ultrasound images using a hierarchical Bayesian model based on generalized Gaussian priors," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3736–3750, 2016.

[7] J. A. Jensen, "Deconvolution of ultrasound images," *Ultrasonic imaging*, vol. 14, no. 1, pp. 1–15, 1992.

[8] O. Michailovich and A. Tannenbaum, "Blind deconvolution of medical ultrasound images: A parametric inverse filtering approach," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 3005–3019, 2007.

[9] H. Ayasso and A. Mohammad-Djafari, "Joint NDT image restoration and segmentation using Gauss–Markov–Potts prior models and variational Bayesian computation," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2265–2277, 2010.

[10] A. Pirayre, Y. Zheng, L. Duval, and J.-C. Pesquet, "HOGMep: Variational Bayes and higher-order graphical models applied to joint image recovery and segmentation," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3775–3779.

[11] M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tourneret, A. O. Hero, and S. McLaughlin, "A survey of stochastic simulation and optimization methods in signal processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 2, pp. 224–241, 2016.

[12] R. M. Neal, "MCMC using Hamiltonian dynamics," *Handbook of Markov Chain Monte Carlo*, vol. 2, no. 11, pp. 2, 2011.

[13] C. P. Robert, V. Elvira, N. Tawn, and C. Wu, "Accelerating MCMC algorithms," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 10, no. 5, 2018.

[14] A. Durmus, E. Moulines, and M. Pereyra, "Efficient Bayesian computation by proximal Markov chain Monte Carlo: when Langevin meets Moreau," *SIAM Journal on Imaging Sciences*, vol. 11, no. 1, pp. 473–506, 2018.

[15] M. Pereyra, "Proximal Markov chain Monte Carlo algorithms," *Statistics and Computing*, vol. 26, no. 4, pp. 745–760, 2016.

[16] A. Schreck, G. Fort, S. Le Corff, and E. Moulines, "A shrinkage-thresholding Metropolis adjusted Langevin algorithm for Bayesian variable selection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 2, pp. 366–375, 2016.

[17] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, 2017.

[18] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212. Springer, 2011.

[19] A. M. Stuart, J. Voss, and P. Wilberg, "Conditional path sampling of SDEs and the Langevin MCMC method," *Communications in Mathematical Sciences*, vol. 2, no. 4, pp. 685–697, 2004.

[20] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "Variable metric forward–backward algorithm for minimizing the sum of a differentiable function and a convex function," *Journal of Optimization Theory and Applications*, vol. 162, no. 1, pp. 107–132, 2014.

[2] Following Landau's notation, we will write that $F(u) = o(\|u-x\|)$, where $F : \mathbb{R}^n \to \mathbb{R}$ and $x \in \mathbb{R}^n$, if $F(u)/\|u-x\| \to 0$ as $u \to x$.

[21] P. L. Combettes, D. Dũng, and B. C. Vũ, "Proximity for sums of composite functions," *Journal of Mathematical Analysis and applications*, vol. 380, no. 2, pp. 680–688, 2011.

[22] Y. Marnissi, E. Chouzenoux, A. Benazza-Benyahia, and J.-C. Pesquet, "Majorize-minimize adapted Metropolis-Hastings algorithm," *HAL preprint HAL:01909153*, 2018.

[23] G. O. Roberts and O. Stramer, "Langevin diffusions and Metropolis-Hastings algorithms," *Methodology and computing in applied probability*, vol. 4, no. 4, pp. 337–357, 2002.

[24] E. D. Schifano, R. L. Strawderman, and M. T. Wells, "Majorization-minimization algorithms for nonsmoothly penalized objective functions," *Electronic Journal of Statistics*, vol. 4, pp. 1258–1299, 2010.

[25] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame-based inverse problems," *Inverse Problems*, vol. 23, no. 4, pp. 1495–1518, June 2007.

[26] J. A. Jensen, "Simulation of advanced ultrasound systems using Field II," in *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2004, pp. 636–639.

[27] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[28] S. Mallat, *A wavelet tour of signal processing*, Elsevier, 1999.

[29] X. Cai, R. Chan, M. Nikolova, and T. Zeng, "A three-stage approach for segmenting degraded color images: Smoothing, lifting and thresholding (SLaT)," *Journal of Scientific Computing*, vol. 72, no. 3, pp. 1313–1332, 2017.

[30] S. Becker and J. Fadili, "A quasi-Newton proximal splitting method," in *Advances in Neural Information Processing Systems*, 2012, pp. 2618–2626.

[31] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[32] S. Krishnan, K. W. Rigby, and M. O'donnell, "Improved estimation of phase aberration profiles," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 44, no. 3, pp. 701–713, 1997.

[33] P. L. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[34] N. Pustelnik and L. Condat, "Proximity operator of a sum of functions; application to depth map estimation," *IEEE Signal Processing Letters*, vol. 24, no. 12, pp. 1827–1831, 2017.