Bayesian anti-sparse coding – Complementary results and supporting materials

Clément Elvira, *Student Member, IEEE*, Pierre Chainais, *Senior Member, IEEE* and Nicolas Dobigeon, *Senior Member, IEEE*

This document contains supplementary material for [36]. Section I presents and compares two other strategies to generates samples according to the so called democratic distribution. Section II describes an additional experiments that assess the correctness of the proposed samplers, following [40]. Section III is a more detailed version of Section B in [36]. Section IV details the computation of the two first moments of the democratic distribution. Section V is dedicated to the proof of the assertion in Remark 2 of [36] concerning the convergence in law of a rescale marginal. Section VI is a detailed proof of Property 3 in [36]. Section VII briefly presents the double-sided Gamma distribution.

I. Other strategies to sample the democratic distribution

1) Gibbs sampler-based random generator: Property 4 can be exploited to design a democratic random variate generator through the use of a Gibbs sampling scheme. It consists of successively drawing the components x_n according to the conditional distributions (18), defined as the mixtures of uniform and truncated Laplacian distributions. After a given number T_{bi} of burn-in iterations, this generator, described in Algo. 1, provides samples asymptotically distributed according to the democratic distribution $\mathcal{D}_N(\lambda)$.

Algorithm 1: Democratic random variate generator using a Gibbs sampling scheme.

Input: Parameter λ, dimension N, number of burn-in iterations T_{bi}, total number of iterations T_{MC}, initialization $\mathbf{x}^{(1,0)} = \left[x_1^{(1,0)}, \ldots, x_N^{(1,0)}\right]^T_{\mathrm{MC}}$

for $t \leftarrow 1$ **to** T_{MC} **do**
  for $n \leftarrow 1$ **to** N **do**
    Set $\mathbf{x}_{\backslash n}^{(t,n-1)} = \left[x_1^{(t,n)}, \ldots, x_{n-1}^{(t,n)}, x_{n+1}^{(t,n-1)}, \ldots, x_N^{(t,n-1)}\right]^T$;
    Draw $x_n^{(t,n)}$ according to (18);
    Set $\mathbf{x}^{(t,n)} = \left[x_1^{(t,n)}, \ldots, x_{n-1}^{(t,n)}, x_n^{(t,n)}, x_{n+1}^{(t,n-1)}, \ldots, x_N^{(t-1,n-1)}\right]^T$;
  end
  Set $\mathbf{x}^{(t+1,0)} = \mathbf{x}^{(t,N)}$;
end

Output: $\mathbf{x}^{(t,0)} \sim \mathcal{D}_N(\lambda)$ (for $t > T_{\mathrm{bi}}$)

2) P-MALA-based random generator: An alternative to draw samples according to the democratic distribution is the proximal Metropolis-adjusted Langevin algorithm (P-MALA) introduced in [35]. P-MALA builds a Markov chain $\left\{\mathbf{x}^{(t)}\right\}_{t=1}^{T_{\mathrm{MC}}}$ whose stationary distribution is of the form

$$\mathrm{p}(\mathbf{x}) \propto \exp(-g(\mathbf{x}))$$

where g is a positive convex function with $\lim\limits_{\|\mathbf{x}\| \to \infty} g(\mathbf{x}) = +\infty$. It relies on successive Metropolis Hastings moves with Gaussian proposal distributions whose mean has been chosen as the proximal operator of $g(\cdot)$ evaluated at the current state of the chain. In the particular case of the democratic distribution, P-MALA can be implemented by exploiting the derivations

Clément Elvira and Pierre Chainais are with Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL - Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France (e-mail: {Clement.Elvira, Pierre.Chainais}@ec-lille.fr).

Nicolas Dobigeon is with the University of Toulouse, IRIT/INP-ENSEEIHT, CNRS, 2 rue Charles Camichel, BP 7122, 31071 Toulouse cedex 7, France (e-mail: Nicolas.Dobigeon@enseeiht.fr).

Part of this work has been funded thanks to the BNPSI ANR project no. ANR-13-BS-03-0006-01.

Part of this work was presented during IEEE SSP Workshop 2016 [1].

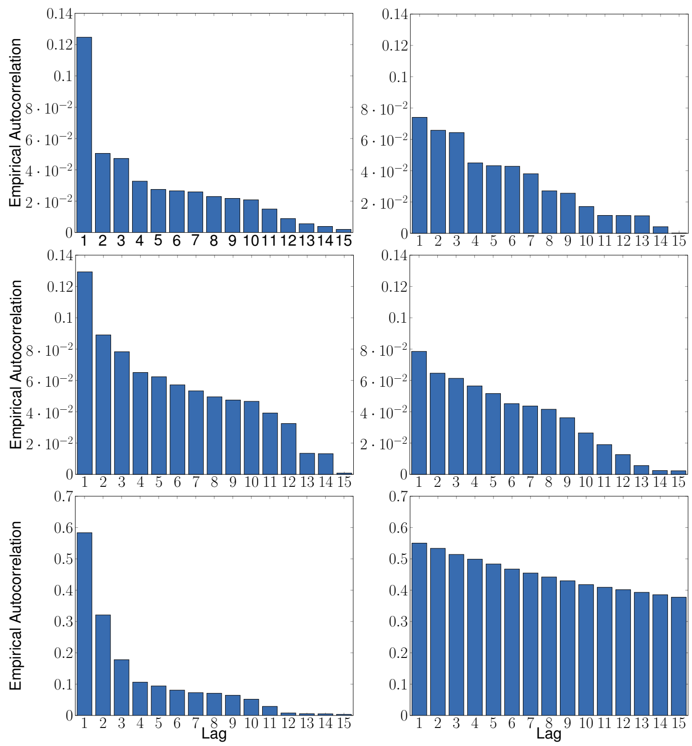

Fig. 1. First 15 lags of the empirical autocorrelation function when generating 500 samples drawn from $\mathcal{D}_N(\lambda)$ using the exact (top), Gibbs sampler-based (middle) and P-MALA (bottom) variate generators with $\lambda = 3$ for $N = 2$ (left) and $N = 50$ (right).

in Section II-E, where $g(\cdot) = g_1(\cdot)$ has been defined in (20). More precisely, at iteration t of the sampler, a candidate $\mathbf{x}^*$ is proposed as

$$\mathbf{x}^*|\mathbf{x}^{(t-1)} \sim \mathcal{N}\left(\operatorname{prox}_{g_1}^{\delta/2}\left(\mathbf{x}^{(t-1)}\right), \delta \mathbf{I}_N\right). \tag{S.1}$$

Then this candidate is accepted as the new state $\mathbf{x}^{(t)}$ with probability

$$\alpha = \min\left(1, \frac{\mathrm{p}\left(\mathbf{x}^*|\lambda\right)}{\mathrm{p}\left(\mathbf{x}^{(t-1)}|\lambda\right)} \frac{q\left(\mathbf{x}^{(t-1)}|\mathbf{x}^*\right)}{\mathrm{q}\left(\mathbf{x}^*|\mathbf{x}^{(t-1)}\right)}\right) \tag{S.2}$$

where $\mathrm{q}\left(\mathbf{x}^*|\mathbf{x}^{(t-1)}\right)$ is the pdf of the Gaussian distribution given by (S.1). The algorithmic parameter δ is empirically chosen such that the acceptance rate of the sampler lies between 0.4 and 0.6. The full algorithmic scheme is available in Algo. 2.

Algorithm 2: Democratic random variate generator using P-MALA.

Input: Parameter λ, dimension N, number of burn-in iterations T_{bi}, total number of iterations T_{MC}, algorithmic parameter δ, initialization $\mathbf{x}^{(0)}$

for $t \leftarrow 1$ **to** T_{MC} **do**
- Draw $\mathbf{x}^*|\mathbf{x}^{(t-1)} \sim \mathcal{N}\left(\operatorname{prox}_{g_1}^{\delta/2}\left(\mathbf{x}^{(t-1)}\right), \delta \mathbf{I}_N\right)$;
- Compute α following (S.2) ;
- Draw $w \sim \mathcal{U}(0,1)$;
- **if** $w < \alpha$ **then**
  - Set $\mathbf{x}^{(t)} = \mathbf{x}^*$;
- **else**
  - Set $\mathbf{x}^{(t)} = \mathbf{x}^{(t-1)}$;
- **end**

end

Output: $\mathbf{x}^{(t)} \sim \mathcal{D}_N(\lambda)$ (for $t > T_{\mathrm{bi}}$)

TABLE I

CUMULATIVE TIMES TO DRAW 500 SAMPLES ACCORDING THE DEMOCRATIC DISTRIBUTION $\mathcal{D}_N(\lambda)$ WITH $\lambda = 3$ FOR VARIOUS N.

N	Exact (ms)	Gibbs (ms)	P-MALA (ms)
2	2.06×10^{-1}	2.05×10^{3}	1.11×10^{2}
5	3.30×10^{-1}	2.66×10^{3}	1.12×10^{2}
10	3.46×10^{-1}	3.60×10^{3}	1.14×10^{2}
50	9.24×10^{-1}	1.01×10^{4}	1.27×10^{2}
100	1.71×10^{0}	1.85×10^{4}	1.46×10^{2}

3) Random generator performance comparison: Figure 1 compares the first 15 lags of the empirical autocorrelation function (ACF), computed with 500 samples drawn from the democratic distribution $\mathcal{D}_N(3)$ for $N = 2$ (left) and $N = 50$ (right) using the exact (top), Gibbs sampler-based (middle) and P-MALA (bottom) variate generators. In lower dimensional cases, the chain generated with exact sampling has remarkably lower autocorrelation.

Computational times required to generate 1000 samples from the democratic distribution for various dimensions are reported in Table I. These results show that the Gibbs sampler-based method has a significantly higher cost when compared with the exact random generator. Whereas the exact sampler easily scales in higher dimension, the Gibbs based sampler needs approximately 10^4x more time. This is explained by its intrinsic algorithmic structure: the Gibbs sampler-based method requires to draw a multinomial variable for each component, followed by either exponential or uniform distributed variables. Conversely, exact random generator only needs to generate one gamma distributed variable and $(N-1)$ uniform samples.

From these findings, the Gibbs sampler-based and P-MALA strategies might seem out of interest since significantly outperformed by the exact random generator in terms of time computation and mixing performance. However, both exhibit interesting properties that can be exploited in a more general scheme. First, the Gibbs sampler-based generator shows that each component of a democratically distributed vector can be easily generated conditionally on the others. Then, P-MALA exploits the algorithmic derivation of the proximity operator associated with $g_1(\cdot)$ to draw vectors asymptotically distributed according to the democratic distribution. This opens the door to extended schemes for sampling according to a posterior distribution resulting from a democratic prior when possibly no exact sampler is available, see Section III.

II. ADDITIONAL EXPERIMENT : ASSESSMENT OF THE BAYESIAN ANTI-SPARSE CODING ALGORITHMS

We first consider an experiment to assess the validity of the Monte Carlo algorithms introduced in Section III-B and, in particular, the two methods proposed to sample according to the conditional distribution of the representation vector $f\left(\mathbf{x}|\mathbf{y}, \sigma^2, \mu\right)$, see step 7 in Algo. 2. To this aim, the so-called *successive conditional sampling* strategy proposed by Geweke in [40] is followed for fixed nuisance parameters σ^2 and μ. This procedure does not fully assert the correctness of the sampler but it may help to detect errors, e.g., as in [42]. More precisely, it consists of drawing a sequence $\left\{\mathbf{x}^{(t)}, \mathbf{y}^{(t)}\right\}_{t=1}^{T_{\mathrm{MC}}}$ asymptotically distributed according to the joint distribution $f\left(\mathbf{x}, \mathbf{y}|\sigma^2, \mu\right)$ using the Gibbs sampler described in Algo. 3.

Algorithm 3: Successive conditional sampling

Input: Residual variance σ^2, democratic parameter μ, coding matrix $\mathbf{H}$.

Sample $\mathbf{x}^{(0)}$ according to $\mathcal{D}_N(\mu N)$;

for $t \leftarrow 1$ **to** T_{MC} **do**

  Sample $\mathbf{y}^{(t)}|\mathbf{x}^{(t-1)}, \sigma^2 \sim \mathcal{N}(\mathbf{H}\,\mathbf{x}^{(t)}, \sigma^2)$;

  Sample $\mathbf{x}^{(t)}|\mathbf{y}^{(t)}, \mu, \sigma^2$ using, either

  - Gibbs steps, i.e., following (36) ;

  Or

  - P-MALA step, i.e., following (40) and (41);

end

Output: A collection of samples $\left\{\mathbf{y}^{(t)}, \mathbf{x}^{(t)}\right\}_{t=T_{\mathrm{bi}}+1}^{T_{\mathrm{MC}}}$ asymptotically distributed according to $f\left(\mathbf{y}, \mathbf{x}|\sigma^2, \mu\right)$

One can notice that this algorithm boils down to successively sample according to the Gaussian likelihood distribution $f\left(\mathbf{y}|\mathbf{x}, \sigma^2, \mu\right)$ and the conditional posterior of interest $f\left(\mathbf{x}|\mathbf{y}, \sigma^2, \mu\right)$. This later step is achieved using either the component-wise Gibbs sampler or P-MALA technique described in Section III-B3. Within this framework, the generated samples $\mathbf{x}^{(t)}$ should be asymptotically distributed according to the prior democratic distribution $f\left(\mathbf{x}|\mu\right)$. Thus they can be exploited to specifically assess the validity of this step, by resorting to the properties of this distribution, see Section II. In this experiment, we propose to focus on one of these properties: the absolute value of the dominant component $f\left(|x_n|\middle|\mu, x_n \in \mathcal{C}_n\right)$ follows the gamma distribution $\mathcal{G}(N, \lambda)$ with $\lambda = N\mu$, see (13). Figure 2 (left) compares the theoretical pdf of this gamma distribution

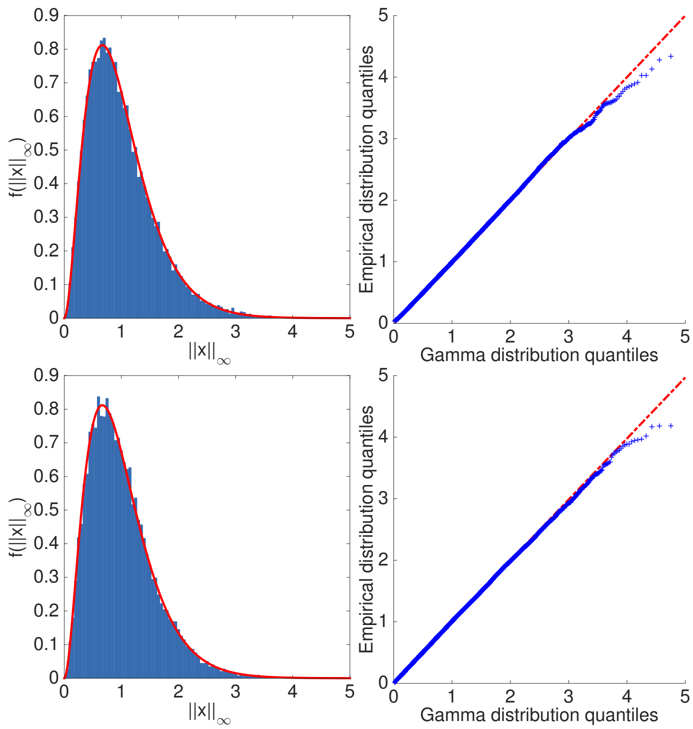

Fig. 2. Left: theoretical Gamma pdf of the dominant component absolute value (red) and empirical pdf (blue) resulting from 2×10^4 samples generated by successive conditional sampling (Algo. 3) with the Gibbs sampler (top) and P-MALA (bottom). Right: corresponding Q-Q plots.

with the empirical pdfs computed from $T_{\mathrm{MC}} = 2 \times 10^4$ samples generated by the Geweke's scheme with the Gibbs (top) and P-MALA (bottom) samplers, where $M = N = 3$, $\mu = 2$ and $\sigma^2 = 0.25$. Figure 2 (right) shows the corresponding quantile-quantile (Q-Q) plots which tends to ascertain the validity of the two versions of the MCMC algorithm.

III. Marginal distributions

A. Any marginal distributions

This appendix derives the marginal distribution of any subset, exhibited by Lemma 2. Let $\mathbf{x} = [x_1, \ldots, x_N]^T$ be a random vector drawn from the democratic distribution $\mathcal{D}_N(\lambda)$. For any positive integer $J < N$, let $\mathcal{K}_J$ be a J-element subset of $\{1, \ldots, N\}$ and $\mathbf{x}_{\backslash \mathcal{K}_J}$ the sub-vector of $\mathbf{x}$ whose J elements indexed by $\mathcal{K}_J$ have been removed. To obtain the marginal distribution stated in Lemma 2, one have have to sum

$$\mathrm{p}\left(\mathbf{x}_{\backslash \mathcal{K}_J}\right) = \frac{2^J}{C_N(\lambda)} \int_{\mathbb{R}_+^J} \exp\left(-\lambda \left\|\mathbf{x}\right\|_\infty\right) \mathrm{d}\, x_j^{\otimes j \in \mathcal{K}_J} \tag{S.3}$$

for any J and any subset $\mathcal{K}_J$.

The proof is done by induction. To that aim, let us consider the following assertion, indexed by J and denoted $\mathcal{P}(J)$

$\mathcal{P}(J)$: *"For any J-element subset $\mathcal{K}_J$ of $\{1, \ldots, N\}$, the marginal distribution given in Lemma 2 holds".*

a) Initialization: for $J = 0$ the subset $\mathcal{K}_J$ is empty and the marginal distribution should be nothing more than the pdf of the democratic distribution $\mathcal{D}_N(\lambda)$. Since it is verified, $\mathcal{P}(0)$ is true and the recursion is initialized.

b) Induction: let J be an integer of $\{1, \ldots, N-1\}$, and suppose $\mathcal{P}(J)$ true. Let k be any integer of $\{1, \ldots, N\} \backslash \mathcal{K}_J$ and consider the set $\mathcal{K}_{J+1} = \mathcal{K}_J \cup \{k\}$. Since $\mathcal{P}(J)$ holds, the marginal distribution $\mathrm{p}\left(\mathbf{x}_{\backslash \mathcal{K}_{J+1}}\right)$ can be computed as follows

$$\begin{aligned}
\mathrm{p}\left(\mathbf{x}_{\backslash \mathcal{K}_{J+1}}\right) =& \int_{\mathbb{R}} \mathrm{p}\left(\mathbf{x}_{\backslash \mathcal{K}_J}\right) \mathrm{d}x_k \\
=& 2 \int_{\mathbb{R}_+} \frac{2^J}{C_N(\lambda)} \sum_{j=0}^{J} \left(\binom{J}{j} \frac{(J-j)!}{\lambda^{J-j}} \left\|\mathbf{x}_{\backslash \mathcal{K}_J}\right\|_\infty^j \right) \times \exp\left(-\lambda \left\|\mathbf{x}_{\backslash \mathcal{K}_J}\right\|_\infty\right) \mathrm{d}x_k .
\end{aligned}$$

Since all functions in the sum are "integrable", the finite sum and the integral can be inverted, leading to the integration of $J+1$ similar functions. Constant apart, for $j=0$, $\mathbb{R}_+$ is partitioned as follows $\mathbb{R}_+ = \left[0, \left\|\mathbf{x}_{\setminus\mathcal{K}_{J+1}}\right\|_\infty\right) \cup \left[\mathbf{x}_{\setminus\mathcal{K}_{J+1}}, +\infty\right)$ leading to

$$\int_{\mathbb{R}_+} \exp\left(-\lambda \left\|\mathbf{x}_{\setminus\mathcal{K}_J}\right\|_\infty\right) \mathrm{d}x_k = \left\|\mathbf{x}_{\setminus\mathcal{K}_{J+1}}\right\|_\infty \exp(-\lambda \left\|\mathbf{x}_{\setminus\mathcal{K}_{J+1}}\right\|_\infty) + \frac{1}{\lambda}\exp\left(-\lambda \left\|\mathbf{x}_{\setminus\mathcal{K}_{J+1}}\right\|_\infty\right) \tag{S.4}$$

The procedure is the same for $j>0$ but requires one or several integrations by parts. The integrator by parts is comes from the integration over $\left[\left\|\mathbf{x}_{\setminus\mathcal{K}_{J+1}}\right\|_\infty, +\infty\right)$, leading to

$$\begin{aligned}\int_{\left\|\mathbf{x}_{\setminus\mathcal{K}_{J+1}}\right\|_\infty}^{+\infty} & {x_k}^j e^{-\lambda x_k} \mathrm{d}x_k = \\ & \left\|\mathbf{x}_{\setminus\mathcal{K}_{J+1}}\right\|_\infty^{j+1} e^{-\lambda\left\|\mathbf{x}_{\setminus\mathcal{K}_{J+1}}\right\|_\infty} + \sum_{l=1}^{j+1} \frac{j!}{(j+1-l)!}\frac{1}{\lambda^l} \left\|\mathbf{x}_{\setminus\mathcal{K}_{J+1}}\right\|_\infty^{j+1-l} e^{-\lambda\left\|\mathbf{x}_{\setminus\mathcal{K}_{J+1}}\right\|_\infty}.\end{aligned} \tag{S.5}$$

Note that the last equation also holds for $j=0$, leading to

$$\begin{aligned}\mathrm{p}\left(\mathbf{x}_{\setminus\mathcal{K}_{J+1}}\right) = & \frac{2^{J+1}}{C_N(\lambda)} e^{-\lambda\left\|\mathbf{x}_{\setminus\mathcal{K}_{J+1}}\right\|_\infty} \\ & \times \sum_{j=0}^{J} \binom{J}{j} \frac{(J-j)!}{\lambda^{J-j}} \left(\left\|\mathbf{x}_{\setminus\mathcal{K}_{J+1}}\right\|_\infty^{j+1} + \sum_{l=1}^{j+1} \frac{j!}{(j+1-l)!}\frac{1}{\lambda^l} \left\|\mathbf{x}_{\setminus\mathcal{K}_{J+1}}\right\|_\infty^{j+1-l} \right).\end{aligned}$$

Hence, the whole expression can be factorized as a product of two functions of $\left\|\mathbf{x}_{\setminus\mathcal{K}_{J+1}}\right\|_\infty$, whose second term is polynomial. The last step to state that $\mathcal{P}(J+1)$ is true is to gather the terms with the same degree. To that, one need to distinguish tree cases

c) Degree $J+1$: there is only one terms of degree $J+1$, coming from the term out of the sum for $j=J$, whose value is

$$\binom{J}{J}\frac{(J-J)!}{\lambda^{J-J}} = 1 = \binom{J+1}{J+1}\frac{(J+1-J-1)!}{\lambda^{J+1-J-1}}. \tag{S.6}$$

d) Degree 0: The terms of degree 0 appears for all value of j where $l=j+1$. Hence its value is

$$\begin{aligned}\sum_{j=0}^{J} \binom{J}{j} \frac{(J-j)!}{\lambda^{J-j}} \times \frac{j!}{\lambda^{j+1}} =& \frac{1}{\lambda^{J+1}} \sum_{j=0}^{J} \binom{J}{j} (J-j)!j! \\ =& \binom{J+1}{0} \frac{(J+1)!}{\lambda^{J+1}}.\end{aligned} \tag{S.7}$$

e) Degree $0 < p < N+1$: The terms of degree 0 comes from one term out of the sum, and all $j > p$ such that $l = j+1-p$. Hence

$$\begin{aligned}\binom{J}{p-1}\frac{(J-p+1)!}{\lambda^{J-p+1}} + \sum_{j=p}^{J}\binom{J}{j}\frac{(J-j)!}{\lambda^{J-j}}\frac{j!}{p!}\frac{1}{\lambda^{j+1-p}} & \\ =& \frac{1}{\lambda^{J+1-p}} \left(\binom{J}{p-1}(J-p+1)! + \frac{1}{p!}\sum_{j=p}^{J}\binom{J}{j}(J-j)!j! \right) \\ =& \frac{(J+1-p)!}{\lambda^{J+1-p}} \left(\binom{J}{p-1} + \frac{J!}{p!(J-p)!} \right) \\ =& \frac{(J+1-p)!}{\lambda^{J+1-p}} \binom{J+1}{p}.\end{aligned} \tag{S.8}$$

Therefore $\mathcal{P}(j+1)$ is true and Lemma 2 holds by induction.

B. Two specific distribution

The first specific marginal distribution can be obtained by using Lemma 2 with $J = N-1$ and $\mathcal{K}_J = [1, N] \setminus \{n\}$. Thus $\left\|\mathbf{x}_{\setminus\mathcal{K}_J}\right\|_\infty = |x_n|$ and

$$\begin{aligned}\mathrm{p}(x_n) =& \frac{\lambda^N}{2N!} \sum_{j=0}^{N-1} \binom{N-1}{j} \frac{(N-1-j)!}{\lambda^{N-1-j}} |x_n|^j \exp\left(-\lambda |x_n|\right) \\ =& \frac{\lambda}{2N} \sum_{j=0}^{N-1} \frac{1}{j!} \frac{1}{\lambda^{-j}} |x_n|^j \exp\left(-\lambda |x_n|\right),\end{aligned}$$

which the mixture distribution claimed in Equation (9).

The second specific marginal distribution can be obtained by using Lemma 2 with $J = 1$ and $\mathcal{K}_J = \{n\}$. Thus

$$\mathrm{p}\left(\mathbf{x}_{\setminus\mathcal{K}_J}\right) = \frac{\lambda}{2N}\,\frac{2}{C_{N-1}(\lambda)}\left(\frac{1}{\lambda} + \left\|\mathbf{x}_{\setminus\mathcal{K}_J}\right\|_\infty\right)\exp\left(-\lambda|\mathbf{x}_{\setminus\mathcal{K}_J}|\right),$$

which is the distribution claimed in Equation (10).

IV. Two first modes of the democratic distribution

Let $\mathbf{x} = [x_1, \ldots, x_N]^T$ obeying the democratic distribution $\mathcal{D}_N(\lambda)$.

Different approaches are available to compute the mean. The fastest one consists in noticing that the function of interest, $\mathbf{x} \to x\,\mathrm{p}\left(x\right)$ is odd and the integration is symmetric with respect to 0.

To compute the variance, a possible solution consists in using the marginal distribution (9) exhibited in Property 2. Although stated after, the proof of Lemma 2 does not rely on this result. Thus, the marginal distribution of one component is a sum of N double-sided Gamma distributions whose first parameter goes from 1 to N. Hence, using (9) and for $n \in [1, N]$

$$\begin{aligned}
\mathrm{Var}(X_n) &= \int_{\mathbb{R}} x_n^2 \mathrm{p}\left(x_n\right)\mathrm{d}\,x_n \\
&= \frac{1}{N}\sum_{i=1}^{N}\int_{\mathbb{R}} x_n^2\ \mathrm{p}\left(Y_i = x_n | Y_i \sim d\mathcal{G}\left(i,\lambda\right)\right)\mathrm{d}\,x_n \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathrm{Var}(Y_i | Y_i \sim d\mathcal{G}\left(i,\lambda\right)) \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{i(i+1)}{\lambda^2} \\
&= \frac{1}{N\lambda^2}\left(\frac{N(N+1)}{2} + \frac{N(N+1)(2N+1)}{6}\right) \\
&= \frac{(N+1)(N+2)}{3\lambda^2}.
\end{aligned}$$

For more details about the variance of the double-sided Gamma distribution, seeAppendix VII.

For n_1 and n_2 such that $n_1 \neq n_2$, the covariance can be computed following the same scheme, using (9) proposed in Lemma 2

$$\mathrm{Cov}\left[X_{n_1}, X_{n_2}\right] = \int_{\mathbb{R}^2} x_{n_1}x_{n_2}\mathrm{p}\left(x_{n_1}, x_{n_2}\right)\mathrm{d}\,x_{n1}\,\mathrm{d}\,x_{n2}. \tag{S.9}$$

Since the marginal distribution $\mathrm{p}\left(X_{n_1}, X_{n_2}\right)$ is expressed in terms of $\|\ \|_\infty$, the integration domain will again be split into cones, leading to

$$\begin{aligned}
&\int_{\mathcal{C}_1^+} x_{n_1}x_{n_2}\mathrm{p}\left(x_{n_1}x_{n_2}\right)\mathrm{d}\,x_{n1}\,\mathrm{d}\,x_{n2} \\
&= \int_{x_{n_1}=0}^{+\infty}\int_{x_{n_2}=-|x_{n_1}|}^{|x_{n_1}|} x_{n_1}x_{n_2}\mathrm{p}\left(x_{n_1}x_{n_2}\right)\mathrm{d}\,x_{n_1}\,\mathrm{d}\,x_{n_2} \\
&=\frac{2^{N-2}}{C_N(\lambda)}\sum_{j=0}^{N-2}\frac{\lambda^j}{j!}\int_{x_{n_1}=0}^{+\infty}\int_{x_{n_2}=-x_{n_1}}^{x_{n_1}} x_{n_1}x_{n_2}\left|x_{n_1}\right|^j \times \exp\left(-\lambda\left|x_{n_1}\right|\right)\mathrm{d}\,x_{n_1}\,\mathrm{d}\,x_{n_2} \\
&=\frac{2^{N-2}}{C_N(\lambda)}\sum_{j=0}^{N-2}\frac{\lambda^j}{j!}\int_{0}^{+\infty} x_{n_1}\left|x_{n_1}\right|^j \times \exp\left(-\lambda\left|x_{n_1}\right|\right) \times \underbrace{\left[\frac{1}{2}x_2^2\,\mathrm{d}\,x_{n_2}\right]_{-x_{n_1}}^{x_{n_1}}}_{0}\mathrm{d}\,x_{n_1} \\
&= 0.
\end{aligned}$$

Similarly, the same result is obtained when integrating over $\mathcal{C}_1^-$, $\mathcal{C}_2^+$ and $\mathcal{C}_2^-$.

V. Proof of the convergence of the marginal distribution

The results is formulated following a central limit statement.

Property 1. *Let λ be a positive real number and $\{X_1, \ldots, X_N\}$ be a sequence of random variable, where for any integer N greater than 0 X_N is the first marginal component of a random vector which obeys the democratic distribution $\mathcal{D}_N(\lambda)$. Then, as N approaches infinity, the random variable $\frac{\lambda}{n} X_n$ converge in distribution to a uniform distribution $\mathcal{U}([-1, 1])$.*

Proof. This result can be proven by showing a pointwise convergence of the characteristic function.

Let first N be an integer. As it has been stated in property 2, the random variable associated to the marginal distribution is a mixture of N gamma distributions. By properties of the characteristic function one have

$$\begin{aligned}
\phi_{\frac{\lambda}{N} X_N}(t) =& \phi_{X_N}(\frac{\lambda}{N} t) \\
=& \frac{1}{2N} \sum_{j=1}^{N} \int_{\mathbb{R}} \frac{\lambda^j}{\Gamma(j)} |x_n^{j-1}| \exp\left(-\lambda |x_n| + i \frac{\lambda}{N} t x_n\right) \mathrm{d}x_n \\
=& \frac{1}{2N} \sum_{j=1}^{N} \left(1 - i\frac{t}{N}\right)^{-j} - \frac{1}{2N} \sum_{j=1}^{N} \left(1 + i\frac{t}{N}\right)^{-j} \\
=& \frac{\left(1 - i\frac{t}{N}\right)^{-N} - \left(1 + i\frac{t}{N}\right)^{-N}}{2it}.
\end{aligned}$$

The two last lines were obtained by separating the integration on $\mathbb{R}_+$ and $\mathbb{R}_+$ and recognizing the sum of a geometric sequence. Let us now consider t a real number, and N big enough such that $\frac{t}{N} < 1$. Thus, $\left|i\frac{t}{N}\right| < 1$ and $i\frac{t}{N} \neq 1$ and the logarithm $\log(1 - i\frac{t}{N})$ is defined. The Taylor's theorem gives

$$\log(1 - i\frac{t}{N}) = -i\frac{t}{N} + N\mathcal{O}(\frac{t^2}{N^2}). \tag{S.10}$$

Using this, one have

$$\begin{aligned}
\left(1 - i\frac{t}{N}\right)^{-N} &= \exp\left(-N \log(1 - i\frac{t}{N})\right) \\
&= \exp\left(it + \mathcal{O}(\frac{t^2}{N})\right).
\end{aligned}$$

Similarly, one can show that $\left(1 + i\frac{t}{N}\right)^{-N} = \exp\left(-it + \mathcal{O}(\frac{t^2}{N})\right)$. Thus, when N grows to infinity, the sequence of function $\phi_{\frac{\lambda}{N} X_N}$ converges pointwise to $\frac{e^{it} - e^{it}}{2it}$, which is the characteristic function of the uniform distribution on $[-1, 1]$. □

VI. Probability of a component being dominant

The probability $\mathrm{P}[\mathbf{x} \in \mathcal{C}_n]$ is obtained by computing the volume of the cone $\mathcal{C}_n$. The intrinsic symmetries of the distribution lead to

$$\begin{aligned}
\mathrm{P}[\mathbf{x} \in \mathcal{C}_n] &= \frac{2^N}{C_N(\lambda)} \int_{x_n=0}^{+\infty} \int_{x_{\setminus n}=0}^{x_n} \exp(-\lambda x_n) \mathrm{d}x_1 \ldots \mathrm{d}\, x_N \\
&= \frac{2^N}{C_N(\lambda)} \int_0^{+\infty} x_n^{N-1} \exp(-\lambda x_n) \mathrm{d}\, x_n \\
&= \int_0^{+\infty} \frac{\lambda}{N} \exp(-\lambda x_n) \mathrm{d}\, x_n \\
&= \frac{1}{N}.
\end{aligned}$$

VII. The double-sided gamma distribution

The double-sided Gamma distribution $d\mathcal{G}(a, b)$ is defined as a generalization over $\mathbb{R}$ of the standard Gamma distribution $\mathcal{G}(a, b)$ whose pdf is

$$\forall x \in \mathbb{R}, \quad f_{d\mathcal{G}}(x) = \frac{b^a}{2\Gamma(a)} |x|^{a-1} \exp(-b|x|). \tag{S.11}$$

Although the probability distributions are equals up to a multiplicative constant, the two laws have different moments. It is worth saying

$$\mathbf{E}\left[X\right] = 0 \tag{S.12}$$

$$\mathrm{Var}\left[X\right] = \frac{a(a+1)}{b^2}. \tag{S.13}$$

REFERENCES

[1] C. Elvira, P. Chainais, and N. Dobigeon, "Democratic prior for anti-sparse coding," in *Proc. IEEE-SP Workshop Stat. and Signal Process. (SSP)*, Palma de Mallorca, Spain, June 2016.

[2] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, March 2008.

[3] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed Sensing: Theory and Applications*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge, UK: Cambridge University Press, 2012, ch. 1, pp. 1–64.

[4] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, July 2007.

[5] I. Tosic and P. Frossard, "Dictionary learning," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 27–38, March 2011.

[6] R. Gribonval, "Should penalized least squares regression be interpreted as maximum a posteriori estimation?" *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2405–2410, May 2011.

[7] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[8] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, January 2010.

[9] F. Caron and A. Doucet, "Sparse Bayesian nonparametric regression," in *Proc. Int. Conf. Machine Learning (ICML)*, Helsinki, Finland, July 2008.

[10] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learning Research*, vol. 1, pp. 211–244, 2001.

[11] A. Lee, F. Caron, A. Doucet, and C. Holmes, "A hierarchical Bayesian framework for constructing sparsity-inducing priors," *arXiv.org*, September 2010.

[12] K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "A novel hierarchical Bayesian approach for sparse semisupervised hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 585–599, February 2012.

[13] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, June 2008.

[14] D. Tzikas, A. Likas, and N. Galatsanos, "Variational Bayesian sparse kernel-based blind image deconvolution with Student's-t priors," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 753–764, April 2009.

[15] C. Févotte and S. J. Godsill, "A Bayesian approach for blind separation of sparse sources," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 6, pp. 2174–2188, November 2006.

[16] N. Dobigeon, A. O. Hero, and J.-Y. Tourneret, "Hierarchical Bayesian sparse image reconstruction with application to MRFM," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2059–2070, September 2009.

[17] C. Soussen, J. Idier, D. Brie, and J. Duan, "From Bernoulli-Gaussian deconvolution to sparse signal restoration," *IEEE Trans. Signal Process.*, vol. 29, no. 10, pp. 4572–4584, October 2011.

[18] L. Chaari, H. Batatia, N. Dobigeon, and J.-Y. Tourneret, "A hierarchical sparsity-smoothness Bayesian model for $\ell_0 - \ell_1 - \ell_2$ regularization," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 1901–1905.

[19] E. Remes, "Sur une propriété extrémale des polynômes de Tchebychef," *Communications de l'Institut des Sciences Mathématiques et Mécaniques de l'Université de Kharkoff et de la Société Mathématique de Kharkoff*, vol. 13, no. 1, pp. 93–95, 1936.

[20] T. W. Parks and J. H. McClellan, "Chebyshev approximation for nonrecursive digital filters with linear phase," *IEEE Trans. Circ. Theory*, vol. 19, no. 2, pp. 189–194, March 1972.

[21] J. H. McClellan and T. W. Parks, "A personal history of the Parks-McClellan algorithm," *IEEE Signal Process. Mag.*, vol. 22, no. 2, pp. 82–86, March 2005.

[22] L. W. Neustadt, "Minimum effort control systems," *J. SIAM Control*, vol. 1, no. 1, pp. 16–31, 1962.

[23] J. A. Cadzow, "Algorithm for the minimum-effort problem," *IEEE Trans. Autom. Contr.*, vol. 16, no. 1, pp. 60–63, 1971.

[24] Y. Lyubarskii and R. Vershynin, "Uncertainty principles and vector quantization," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3491–3501, July 2010.

[25] Z. Cvetković, "Resilience properties of redundant expansions under additive noise and quantization," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 644–656, March 2003.

[26] A. R. Calderbank and I. Daubechies, "The pros and cons of democracy," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1721–1725, June 2002.

[27] B. Farrell and P. Jung, "A Kashin approach to the capacity of the discrete amplitude constrained Gaussian channel," in *Proc. Int. Conf. Sampling Theory and Applications (SAMPTA)*, Marseille, France, May 2009.

[28] J. Ilic and T. Strohmer, "PAPR reduction in OFDM using Kashin's representation," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Comm.*, Perugia, Italy, 2009, pp. 444–448.

[29] C. Studer, Y. Wotao, and R. G. Baraniuk, "Signal representations with minimum ℓ_∞-norm," in *Proc. Ann. Allerton Conf. Comm. Control Comput. (Allerton)*, 2012, pp. 1270–1277.

[30] C. Studer, T. Goldstein, W. Yin, and R. G. Baraniuk, "Democratic representations," *IEEE Trans. Inf. Theory*, 2014. [Online]. Available: http://arxiv.org/abs/1401.3420/

[31] J.-J. Fuchs, "Spread representations," in *Proc. IEEE Asilomar Conf. Signals, Systems, Computers*, 2011.

[32] H. Jegou, T. Furon, and J.-J. Fuchs, "Anti-sparse coding for approximate nearest neighbor search," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, 2012, pp. 2029–2032.

[33] C. Studer and E. G. Larsson, "PAR-aware large-scale multi-user MIMO-OFDM downlink," *IEEE J. Sel. Areas Comm.*, vol. 31, no. 2, pp. 303–313, February 2013.

[34] J. Tan, D. Baron, and L. Dai, "Wiener filters in Gaussian mixture signal estimation with ℓ_∞-norm error," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6626–6635, October 2014.

[35] M. Pereyra, "Proximal markov chain monte carlo algorithms," *Statistics and Computing*, vol. 26, no. 4, pp. 745–760, 2016.

[36] C. Elvira, P. Chainais, and N. Dobigeon, "Bayesian anti-sparse coding," 2016, submitted.

[37] P. Kabal, "Quantizers for the gamma distribution and other symmetrical distributions," *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, vol. 32, no. 4, pp. 836–841, August 1984.

[38] J.-J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *CR Acad. Sci. Paris Sér. A Math*, vol. 255, pp. 2897–2899, 1962.

[39] L. Condat, "Fast projection onto the simplex and the l1 ball," *Mathematical Programming*, pp. 1–11, 2015.

[40] J. Geweke, "Getting it right: Joint distribution tests of posterior simulators," *J. Amer. Stat. Assoc.*, vol. 99, pp. 799–804, September 2004.

[41] N. Chopin, "Fast simulation of truncated Gaussian distributions," *Statistics and Computing*, vol. 21, no. 2, pp. 275–288, 2010.

[42] D. Knowles and Z. Ghahramani, "Nonparametric bayesian sparse factor models with application to gene expression modeling," *Ann. Appl. Stat.*, vol. 5, no. 2B, pp. 1534–1552, June 2011.

[43] R. Balu, T. Furon, and H. Jégou, “Beyond ”project and sign” for cosine estimation with binary codes,” in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, May 2014, pp. 68 884–6888.