DEMOCRATIC PRIOR FOR ANTI-SPARSE CODING

Clément Elvira[(1,*)], *Pierre Chainais*[(1)] *and Nicolas Dobigeon*[(2)]

[(1)] Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL, F-59000 Lille, France
[(2)] University of Toulouse, IRIT/INP-ENSEEIHT, 31071 Toulouse cedex 7, France

ABSTRACT

Anti-sparse coding aims at spreading the information uniformly over representation coefficients and can be naturally expressed through an ℓ_∞-norm regularization. This paper derives a probabilistic formulation of such a problem. A new probability distribution is introduced.This so-called *democratic* distribution is then used as a prior to promote anti-sparsity in a linear Gaussian inverse problem. A Gibbs sampler is designed to generate samples asymptotically distributed according to the joint posterior distribution of interest. To scale to higher dimension, a proximal Markov chain Monte Carlo algorithm is proposed as an alternative to Gibbs sampling. Simulations on synthetic data illustrate the performance of the proposed method for anti-sparse coding on a complete dictionary. Results are compared with the recent deterministic variational FITRA algorithm.

Index Terms— Anti-sparse representation, democratic distribution, inverse problem.

1. INTRODUCTION

Sparse representations have been widely advocated for regularizing ill-posed inverse problems. Conversely, spreading the information uniformly over a frame is a desirable property in various applications, e.g., to design robust analog-to-digital conversion schemes [1] or to reduce the peak-to-average power ratio (PAPR) in multi-carrier transmissions [2]. A similar problem has been addressed in [3] where the *Kashin's representation* of a given vector over a tight frame is introduced as the expansion with the smallest possible dynamic range. The underlying optimization problem which consists of minimizing the maximum magnitude of the representation coefficients for an upper-bounded ℓ_2-reconstruction error, have been investigated in depth [4, 5]. In these latest contributions, the optimal expansion is called the *democratic representation*. In [6], the constrained signal representation problems considered in [3] and [5] are converted into their penalized counterpart. More precisely, the so-called *spread* or *anti-sparse representations* result from a variational optimization problem where the admissible range of the coefficients has been penalized through a ℓ_∞-norm

$$\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2\sigma^2}\left\|\mathbf{y}-\mathbf{H}\mathbf{x}\right\|_2^2 + \lambda\left\|\mathbf{x}\right\|_\infty . \tag{1}$$

In (1), $\mathbf{H}$ defines the $M\times N$ coding matrix and σ^2 stands for the variance of the residual resulting from the approximation. Besides, recent applications have capitalized on these latest theoretical and algorithmic advances, including approximate nearest neighbour search [7] and PAPR reduction [8].

*Supported in part by the BNPSI ANR Project ANR-13-BS-03-0006-01.

The present article attempts to derive a Bayesian formulation of the anti-sparse coding problem (1) considered in [6]. Bayesian inference allows fully unsupervised methods to be derived, e.g., by including nuisance parameters and other hyperparameters into the Bayesian model. Moreover, it permits to consider a wide range of Bayesian estimators, beyond the standard penalized-maximum likelihood solution associated with (1). To the best of our knowledge, no such probabilistic anti-sparse representation has been proposed yet. The contributions are threefold. First, a new probability density function (pdf), named *democratic* distribution, is introduced. Then, this pdf is resorted to as a prior distribution in a linear Gaussian inverse problem to build a probabilistic counterpart of the problem in (1), under the maximum a posteriori (MAP) paradigm. Finally, two instances of Markov chain Monte Carlo (MCMC) algorithms are derived to generate samples asymptotically distributed according to the resulting posterior distribution. These samples are subsequently used to approximate various Bayesian estimators.

The paper is organized as follows. Section 2 introduces the *democratic* pdf and its corresponding conditional distributions. For sake of briefness, the proofs and complementary properties associated with this distribution have been omitted but can be found in [9]. Section 3 presents the proposed hierarchical Bayesian model for anti-sparse coding, as well as two inference algorithmic schemes. Section 4 illustrates the performance of the proposed methods on numerical experiments. Conclusions are reported in Section 5.

2. THE DEMOCRATIC DISTRIBUTION

The ℓ_∞-norm penalty evoked in (1) can be used to design a new probability distribution belonging to the exponential family, namely the *democratic distribution*. More precisely, $\mathbf{x}\in\mathbb{R}^N$ is said to be distributed according to the democratic distribution with parameter λ, i.e., $\mathbf{x}\sim\mathcal{D}_N(\lambda)$, if its corresponding pdf is

$$f\left(\mathbf{x}|\lambda\right) = \frac{\lambda^N}{2^N N!}\exp\left(-\lambda\left\|\mathbf{x}\right\|_\infty\right). \tag{2}$$

For illustration, the democratic pdf $\mathcal{D}_2(3)$ is depicted in Fig. 1.

2.1. Conditional distributions

The ℓ_∞-norm in the exponential term implicitly generates a partition of $\mathbb{R}^N$ composed of N double-cones $\mathcal{C}_n$, where the nth component is dominant. More precisely, each cone $\mathcal{C}_n$ is defined by

$$\mathcal{C}_n \triangleq \left\{\mathbf{x}=[x_1,\ldots,x_N]^T\in\mathbb{R}^N \;:\; \forall j\neq n,\ |x_j|<|x_n|\right\}. \tag{3}$$

Intrinsic symmetry properties of the democratic distribution lead to a straightforward equiprobability of having a democratic vector which belongs to any of these cones, i.e.,

$$\mathrm{P}\left[\mathbf{x}\in\mathcal{C}_n\right] \;=\; \frac{1}{N},\ \forall n\in\{1,\ldots,N\}. \tag{4}$$

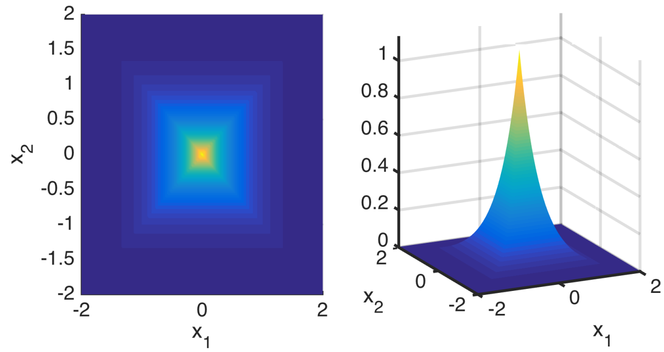

Fig. 1. The democratic pdf $\mathcal{D}_N(\lambda)$ for $N = 2$ and $\lambda = 3$.

As a consequence, conditioning on each $\mathbf{x} \in \mathcal{C}_n$ leads to explicit conditional distributions for its components. More particularly, the conditional distributions of the so-called dominant and non-dominant components of $\mathbf{x} \sim \mathcal{D}_N(\lambda)$ are, respectively,

$$x_n|\mathbf{x}_{\backslash n}, \mathbf{x} \in \mathcal{C}_n \;\sim\; \frac{\lambda}{2} e^{-\lambda\left(|x_n| - \|\mathbf{x}_{\backslash n}\|_\infty\right)} \mathbf{1}_{\mathbb{R}\backslash\mathcal{I}_n}(x_n) \quad (5)$$

$$x_n|\mathbf{x}_{\backslash n}, \mathbf{x} \notin \mathcal{C}_n \;\sim\; \mathcal{U}\left(\mathcal{I}_n\right) \quad (6)$$

where $\mathbf{x}_{\backslash n}$ denotes the vector $\mathbf{x}$ whose nth component has been removed and $\mathcal{I}_n \triangleq \left(-\|\mathbf{x}_{\backslash n}\|_\infty, \|\mathbf{x}_{\backslash n}\|_\infty\right)$. Finally, Eq. (4), (5) and (6) can be used to derive the conditional distribution of one component given the others, by marginalizing out the event that $\mathbf{x}$ belongs to the cone $\mathcal{C}_n$, leading to

$$p\left(x_n|\mathbf{x}_{\backslash n}\right) = (1 - c_n) \frac{1}{2\left\|\mathbf{x}_{\backslash n}\right\|_\infty} \mathbf{1}_{\mathcal{I}_n}(x_n) + c_n \frac{\lambda}{2} e^{-\lambda\left(|x_n| - \|\mathbf{x}_{\backslash n}\|_\infty\right)} \mathbf{1}_{\mathbb{R}\backslash\mathcal{I}_n}(x_n) \quad (7)$$

where
$$c_n \triangleq \mathrm{P}\left[\mathbf{x} \in \mathcal{C}_n | \mathbf{x}_{\backslash n}\right] = \frac{1}{1 + \lambda\left\|\mathbf{x}_{\backslash n}\right\|_\infty}. \quad (8)$$

In other words, the conditional distribution of one component is a mixture of one uniform distribution over $\mathcal{I}_n$ and two shifted exponential distributions over $\mathbb{R}\backslash\mathcal{I}_n$. This result can be exploited to design a random variate generator through the use of a Gibbs sampling scheme [9]. It opens the door to strategies for coefficient-wise sampling according to a posterior distribution resulting from a democratic prior. This will be exploited in Section 3.

2.2. Proximal operator of the negative log-pdf

The democratic pdf can be written as $f(\mathbf{x}) \propto \exp\left(-g(\mathbf{x})\right)$ with $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_\infty$. The proximal operator of g with parameter δ is defined by

$$\mathrm{prox}_g^\delta(\mathbf{x}) = \underset{\mathbf{u}\in\mathbb{R}^N}{\mathrm{argmin}}\; \lambda\|\mathbf{u}\|_\infty + \frac{1}{2\delta}\|\mathbf{x} - \mathbf{u}\|_2^2. \quad (9)$$

Up to authors' knowledge, this minimization does not have any closed-form solution. Nevertheless, the exact solution can be computed with low computational cost, as detailed in [9]. Thus, following the strategy in [10], this proximal operator can be resorted to implement a Monte Carlo algorithm to draw samples from the democratic distribution. This strategy will be also exploited in Section 3 to sample according to a posterior distribution resulting from a democratic prior.

3. BAYESIAN SPARSE CODING

This section describes a Bayesian formulation of the model underlying the problem described by (1).

3.1. Hierarchical Bayesian model

Likelihood Function : Let $\mathbf{y} = [y_1 \ldots y_M]^T$ denote an observed measurement vector. These observations are assumed to be related to an unknown description vector $\mathbf{x} = [x_1 \ldots x_N]^T$ through a known coding matrix $\mathbf{H}$ according to the linear model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}. \quad (10)$$

The residual vector $\mathbf{e} = [e_1 \ldots e_N]^T$ is assumed to be distributed according to the multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}_M, \sigma^2\boldsymbol{I}_M)$. The Gaussian property of the additive residual term yields the following likelihood function

$$f(\mathbf{y}|\mathbf{x}, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{M}{2}} \exp\left[-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2\right]. \quad (11)$$

Description vector prior : The democratic distribution introduced in Section 2 is used as the prior distribution of the N-dimensional unknown vector $\mathbf{x}$ to promote anti-sparsity

$$\mathbf{x} \mid \lambda \sim \mathcal{D}_N(\lambda). \quad (12)$$

In what follows, the hyperparameter λ is set as $\lambda = N\mu$, where μ is assumed to be unknown. This choice allows the hyperparameter to scale with the problem dimension, see [9].

Residual variance prior : A noninformative Jeffreys prior distribution is chosen for the residual variance σ^2

$$f\left(\sigma^2\right) \propto \frac{1}{\sigma^2}. \quad (13)$$

Democratic parameter prior : A conjugate Gamma distribution is chosen as a prior for μ

$$f\left(\mu\right) \propto \mu^a e^{-b\mu} \quad (14)$$

where values of a and b are chosen to obtain a flat prior (e.g., $a = b = 10^{-3}$).

Joint posterior distribution : The likelihood and the priors define above allow the joint posterior distribution to be expressed according to the hierarchical structure

$$f(\mathbf{x}, \sigma^2, \mu|\mathbf{y}) \propto f(\mathbf{y}|\mathbf{x}, \sigma^2) f(\mathbf{x}, \sigma^2|\mu) f(\mu) \quad (15)$$

leading to

$$f(\mathbf{x}, \sigma^2, \mu|\mathbf{y}) \propto \mu^N \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 - \mu N\|\mathbf{x}\|_\infty\right) \times \mu^{a-1}\exp\left(-b\mu\right)\left(\frac{1}{\sigma^2}\right)^{\frac{M}{2}+1}\mathbf{1}_{\mathbb{R}_+}(\sigma^2). \quad (16)$$

Note that for fixed nuisance parameters $\lambda = N\mu$ and σ^2, deriving the MAP estimator associated with (15) is equivalent to solve (1). In an unsupervised framework, these unknown parameters need to be jointly estimated from the measurements or marginalized from the joint posterior, which yields the marginal posterior distribution

$$f\left(\mathbf{x}|\mathbf{y}\right) \propto \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^{-\frac{M}{2}} \left(b + N\|\mathbf{x}\|_\infty\right)^{-(a+N)}. \quad (17)$$

The next paragraph introduces a MCMC algorithm that allows a set of samples $\left\{\mu^{(t)}, \sigma^{2(t)}, \mathbf{x}^{(t)}\right\}_{t=T_{\mathrm{bi}}+1}^{T_{\mathrm{MC}}}$ to be generated according to the posterior distribution (15). Then these samples can be used to approximate the Bayesian estimators, e.g., the minimum mean square error (MMSE) estimator $\hat{\mathbf{x}}_{\mathrm{MMSE}} = \mathrm{E}[\mathbf{x}|\mathbf{y}]$ and the marginal MAP (mMAP) estimator $\hat{\mathbf{x}}_{\mathrm{mMAP}}$ maximizing (17).

3.2. Gibbs sampler

The proposed MCMC algorithm is a Gibbs sampler that consists of successively sampling according to the conditional distributions associated with the joint distribution (15). Its main steps are described in what follows.

Sampling the residual variance : the conditional distribution of the residual variance is the following inverse-Gamma distribution

$$\sigma^2|\mathbf{y},\mathbf{x} \sim \mathcal{IG}\left(\frac{M}{2}, \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2\right). \quad (18)$$

Sampling the democratic hyperparameter : sampling according to $f(\mu|\mathbf{x})$ is achieved as follows

$$\mu|\mathbf{x} \sim \mathcal{G}(a+N, b+N\|\mathbf{x}\|_\infty) \quad (19)$$

Sampling the description vector : The description vector can be sampled component-by-component according to the following 3-mixture of truncated Gaussian distributions

$$x_n|\mathbf{x}_{\setminus n}, \mu, \sigma^2, \mathbf{y} \sim \sum_{i=1}^{3} \omega_{in} \mathcal{N}_{\mathcal{I}_{in}}\left(\mu_{in}, s_n^2\right) \quad (20)$$

where $\mathcal{N}_{\mathcal{I}}(\cdot,\cdot)$ denotes the Gaussian distribution truncated on $\mathcal{I}$ and

$$\mathcal{I}_{1n} = \left(-\infty, -\|\mathbf{x}_{\setminus n}\|_\infty\right),\ \mathcal{I}_{3n} = \left(\|\mathbf{x}_{\setminus n}\|_\infty, +\infty\right),$$
$$\text{and } \mathcal{I}_{2n} = \left(-\|\mathbf{x}_{\setminus n}\|_\infty, \|\mathbf{x}_{\setminus n}\|_\infty\right).$$

The probabilities $\omega_{i,n}$ ($i = 1,\ldots,3$) as well as the means $\mu_{i,n}$ ($i = 1,\ldots,3$) and variances s_n^2 of these truncated Gaussian distributions are given in Appendix A. Sampling according to truncated distributions can be achieved using the strategy proposed in [11].

3.3. Proximal Metropolis Adjusted Langevin Algorithm

The proximal Metropolis adjusted Langevin algorithm (P-MALA) [10] is a scalable alternative to the method presented in section 3.2 to draw the full description vector $\mathbf{x}$. It consists of Metropolis Hastings moves whose proposal distribution is Gaussian with the proximal operator of the negative log-posterior evaluated at the current point as mean. It appears in (16) that the negative log-posterior h is given by

$$h(\mathbf{x}) = \frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_\infty. \quad (21)$$

To the best of our knowledge, no closed forme solution of (21) is available. To alleviate this problem, a first order approximation is considered[1] , as recommended in [10]

$$\text{prox}_h^{\delta/2}(\mathbf{x}) \approx \text{prox}_g^{\delta/2}\left(\mathbf{x} + \delta\,\nabla\left[\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2\right]\right) \quad (22)$$

where $\text{prox}_g^{\delta/2}$ is the proximal operator considered in section 2.2. Hence, at iteration t of the main algorithm, new candidate are proposed according to

$$\mathbf{x}^*|\mathbf{x}^{(t-1)} \sim \mathcal{N}\left(\text{prox}_h^{\delta/2}\left(\mathbf{x}^{(t-1)}\right), \delta\mathbf{I}_N\right) \quad (23)$$

and accepted as a new state $\mathbf{x}^{(t)}$ with probability

$$\alpha = \min\left(1, \frac{f\left(\mathbf{x}^*|\mu,\sigma^2,\mathbf{y}\right)}{f\left(\mathbf{x}^{(t-1)}|\mu,\sigma^2,\mathbf{y}\right)} \frac{q\left(\mathbf{x}^{(t-1)}|\mathbf{x}^*\right)}{q\left(\mathbf{x}^*|\mathbf{x}^{(t-1)}\right)}\right) \quad (24)$$

where q is the pdf of the considered proposal. Following [10], δ is tuned to achieve an acceptance rate between 40% and 60%.

[1] Note that a similar step is involved in the fast iterative truncation algorithm (FITRA) [8], a deterministic counterpart of the proposed algorithm and considered in the next section for comparison.

4. SIMULATION RESULTS ON SYNTHETIC DATA

Performance of the proposed algorithm has been evaluated thanks to numerical experiments on synthetic data. More precisely, anti-sparse codes $\mathbf{x}$ of dimension $N = 50$ are recovered from Gaussian observations $\mathbf{y}$ of size $M = 30$. The $M \times N$ coding matrix $\mathbf{H}$ is generated using randomly subsampled discrete Fourier transform (DFT), since they have shown to yield representations with low ℓ_∞-norm [5]. The mMAP and MMSE estimators discussed in paragraph 3.1 are computed from a total of $T_{\text{MC}} = 12 \times 10^3$ iterations of the two MCMC algorithms, i.e., the full Gibbs sampler and the Gibbs sampler including a P-MALA step, described in paragraph 3.2, including $T_{\text{bi}} = 10 \times 10^3$ burn-in iterations. Performances are evaluated over 50 Monte Carlo simulations and reported in terms of reconstruction error $\text{SNR}_\mathbf{y}$ (to measure the coding quality) and PAPR (to measure anti-sparsity), respectively defined by

$$\text{SNR}_\mathbf{y} = 10\log_{10}\frac{\|\mathbf{y}\|_2^2}{\|y-\mathbf{H}\hat{\mathbf{x}}\|_2^2} \quad (25)$$

$$\text{PAPR} = \frac{N\|\hat{\mathbf{x}}\|_\infty^2}{\|\hat{\mathbf{x}}\|_2^2} \in [1, N] \quad (26)$$

where $\hat{\mathbf{x}}$ refers to an estimator of $\mathbf{x}$.

The two proposed algorithms have been compared with FITRA, a PAPR reduction technique detailed in [8]. FITRA directly solves (1), but in a supervised framework, since it needs the prior knowledge of the two nuisance parameters λ and σ^2, i.e., the product $\beta \triangleq 2\lambda\sigma^2$. Consequently, 3 configurations of FITRA are considered: $\beta = 2\hat{\sigma^2}_{\text{MMSE}}\hat{\lambda}_{\text{MMSE}}$ (FITRA-mmse) where $\hat{\sigma^2}_{\text{MMSE}}$ and $\hat{\lambda}_{\text{MMSE}}$ denote the MMSE estimate of these parameters recovered by the proposed algorithm, and two distinct values of β tuned to reach either a targeted $\text{SNR}_\mathbf{y}$ of 20dB (FITRA-snr) or a targeted PAPR of 1.5 (FITRA-papr), respectively. Finally, the algorithms have been compared with the least-square (LS) solution as well as the MMSE and mMAP estimates resulting from a Bayesian model based on a Gaussian prior in place of the democratic one, to assess the interest of the anti-sparsity promoting prior.

Table 1. $\text{SNR}_\mathbf{y}$ and PAPR for various algorithms. Note that $\text{SNR}_\mathbf{y} > 100$ dB are considered infinite.

	$\text{SNR}_\mathbf{y}$	PAPR
P-MALA mMAP	29.3	2.78
P-MALA MMSE	19.3	3.89
Gibbs mMAP	8.8	3
Gibbs MMSE	4.3	6.9
FITRA-mmse	34.4	1.69
FITRA-papr	12.8	1.5
FITRA-snr	19.9	1.71
LS	∞	6.63
Gibbs mMAP (Gaussian)	∞	5.92
Gibbs MMSE (Gaussian)	73.1	6.79

Table 1 shows the average results for all considered algorithms. Among all the proposed methods, the mMAP estimate obtained with P-MALA has reached in average the highest $\text{SNR}_\mathbf{y}$ (29.3 dB) for the lowest PAPR (2.78). As a comparison, tuning FITRA to reach the same $\text{SNR}_\mathbf{y}$ of 29.3 DB leads to a PAPR of 1.81. Conversely, the full Gibbs sampler leads to solutions with low $\text{SNR}_\mathbf{y}$ and high PAPR.

Fig 2 illustrates the variability of the results in terms of the compromise between PAPR and $\text{SNR}_{\mathbf{y}}$. As a reference the average solutions recovered by FITRA are plotted for a continuous range of the hyperparameter β. Points corresponding to estimates from the three Bayesian methods are also plotted. First one notices the bad results obtained with the full Gibbs sampler which almost always yields solutions with either zero or infinite $\text{SNR}_{\mathbf{y}}$. Further investigations have shown that the non informative priors over the two hyperparameters have not led to a compromise. Then, P- MALA produces solutions close to the critical area determined by FITRA. While outperformed by FITRA for a given PAPR, solutions resulting from the democratic prior have a significantly lower PAPR than solutions resulting from a Gaussian prior, which confirms the interest of the ℓ_∞-penalty.

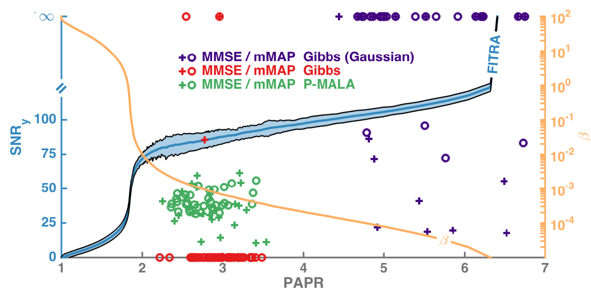

Fig. 2. $\text{SNR}_{\mathbf{y}}$ as a function of PAPR. The blue line is the average results of FITRA for the 50 Monte Carlo Simulations together with a confidence interval.

5. CONCLUSION

This paper introduces a new probability distribution, namely the *democratic distribution*, which is designed to promote anti-sparsity. Once elected as a prior over coefficient, the inference problem was cast as a Bayesian counterpart of anti-sparse coding. A full Gibbs sampler was designed to successively sample in an unsupervised way all parameters according to their individual conditional posterior distributions. An alternative sampler that exploits the proximal operator in a P-MALA step was also proposed. Relevance of the two resulting algorithms was asserted on a synthetic experiment by inferring the representation of a given measurement vector on a known and over-complete dictionary. Performances were compared to other methods : the supervised deterministic PAPR reduction method FITRA, the least-square solution and Bayesian estimators resulting from a Gaussian prior. Despite outperformed by FITRA, the democratic prior distribution was able to promote anti-sparse solutions. The mMAP estimator generally provided more relevant solutions than the MMSE estimator. P-MALA has proposed more satisfying results in terms of $\text{SNR}_{\mathbf{y}}$ and PAPR for a significantly lower computational cost, and the chain has appeared more stable. Future work will investigate the ability of P-MALA to scale to higher dimensions.

A. POSTERIOR DISTRIBUTION OF THE REPRESENTATION COEFFICIENTS

The mean and variances of the truncated Gaussian distributions involved in the mixture distribution (20) are given by

$$\mu_{1n} = \frac{1}{\|\mathbf{h}_n\|^2}\left(\mathbf{h}_n^T\mathbf{e}_n + \sigma^2\lambda\right)$$

$$\mu_{2n} = \frac{1}{\|\mathbf{h}_n\|^2}\left(\mathbf{h}_n^T\mathbf{e}_n\right) \qquad s_n^2 = \frac{\sigma^2}{\|\mathbf{h}_n\|_2^2}$$

$$\mu_{3n} = \frac{1}{\|\mathbf{h}_n\|^2}\left(\mathbf{h}_n^T\mathbf{e}_n - \sigma^2\lambda\right)$$

where $\mathbf{h}_n$ denotes the nth column of $\mathbf{H}$ and $\mathbf{e}_n = \mathbf{y} - \sum_{i\neq n} x_i\mathbf{h}_i$. The weights of each mixture component are

$$\omega_{in} = \frac{u_{in}}{\sum_{j=1}^{3} u_{jn}}$$

with

$$u_{1n} = \exp\left(\frac{\mu_{1n}^2}{2s_n^2} + \lambda\left\|\mathbf{x}_{\setminus n}\right\|_\infty\right)\phi_{\mu_{1n},s_n^2}\left(-\left\|\mathbf{x}_{\setminus n}\right\|_\infty\right)$$

$$u_{2n} = \exp\left(\frac{\mu_{2n}^2}{2s_n^2}\right) \times\left[\phi_{\mu_{2n},s_n^2}\left(\left\|\mathbf{x}_{\setminus n}\right\|_\infty\right) - \phi_{\mu_{2n},s_n^2}\left(-\left\|\mathbf{x}_{\setminus n}\right\|_\infty\right)\right]$$

$$u_{3n} = \exp\left(\frac{\mu_{3n}^2}{2s_n^2} + \lambda\left\|\mathbf{x}_{\setminus n}\right\|_\infty\right)\left(1 - \phi_{\mu_{3n},s_n^2}\left(\left\|\mathbf{x}_{\setminus n}\right\|_\infty\right)\right)$$

where $\phi_{\mu,s^2}(\cdot)$ is the cumulated distribution function of the normal distribution $\mathcal{N}(\mu, s^2)$.

B. REFERENCES

[1] 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.

[2] 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.

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

[4] 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.

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

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

[7] 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.

[8] 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, Feb. 2013.

[9] C. Elvira, P. Chainais, and N. Dobigeon, "Bayesian anti-sparse coding," 2015, submitted. [Online]. Available: http://arxiv.org/abs/1512.06086/

[10] M. Pereyra, "Proximal Markov chain Monte Carlo algorithms," *Stat. Comput.*, May 2015.

[11] N. Chopin, "Fast simulation of truncated gaussian distributions," vol. 21, pp. 275–288, 2011.