

# Bayesian Blind Image Deconvolution using a Hyperbolic-Secant prior

Francisco M. Castro-Macías<sup>a,d,\*</sup>, Fernando Pérez-Bueno<sup>b</sup>, Miguel Vega<sup>c</sup>, Javier Mateos<sup>a</sup>, Rafael Molina<sup>a</sup> and Aggelos K. Katsaggelos<sup>e</sup>

<sup>a</sup> Dpto. Ciencias de la Computación e I. A., Universidad de Granada, Granada, Spain.

<sup>b</sup> Basque Center on Cognition, Brain and Language, Donostia - San Sebastián, Spain

<sup>c</sup> Dpto. Lenguajes y Sistemas Informáticos, Universidad de Granada, Granada, Spain.

<sup>d</sup> Escuela de Doctorado de Ciencias, Tecnologías e Ingenierías, Universidad de Granada, Granada, Spain.

<sup>e</sup> Dept. of Electrical and Computer Engineering, Northwestern University, Evanston (IL), USA.

\* Corresponding author. Email: fcastro@ugr.es

**Abstract**—In this paper we propose the use of the Hyperbolic Secant (HS) distribution as a prior for the Blind Image Deconvolution (BID) problem. It is well-known that when high-pass filters are applied to natural images, the resulting coefficients are sparse. We leverage this property using the HS distribution, a seldom explored Super Gaussian distribution with suitable properties for this problem. Using the Pólya-Gamma distribution, we derive an explicit Gaussian Scale Mixture representation. This representation is then used to propose a novel variational Bayesian algorithm that outperforms state-of-the-art BID methods.

**Index Terms**—Blind Deconvolution, Gaussian Scale Mixture, Hyperbolic Secant, Variational Bayes.

## I. INTRODUCTION

Image deblurring is an important problem in the field of image processing [1]. Its goal is to recover the underlying clean image from an observed blurry and noisy one. Blurriness in the image can be caused by several factors such as camera shake, defocus or atmospheric conditions [2]–[4]. In this work, we focus in those cases where the degradation process can be modelled as

$$\mathbf{y} = \mathbf{Hx} + \boldsymbol{\eta}, \quad (1)$$

where  $\mathbf{y} \in \mathbb{R}^{HW \times 1}$  is the observed blurry image,  $\mathbf{x} \in \mathbb{R}^{HW \times 1}$  is the underlying clean image,  $\mathbf{H} \in \mathbb{R}^{HW \times HW}$  is the convolution matrix obtained from the blur kernel  $\mathbf{h} \in \mathbb{R}^{K \times 1}$ , and  $\boldsymbol{\eta} \in \mathbb{R}^{HW \times 1}$  is additive white Gaussian noise. In the case where the kernel  $\mathbf{h}$  is unknown, recovering  $\mathbf{x}$  from  $\mathbf{y}$  is known as Blind Image Deconvolution (BID). It is a severely ill-posed problem since infinitely many combinations of the blur and *true* image can produce the observed image [5]. To find meaningful solutions, additional restrictions on both  $\mathbf{x}$  and  $\mathbf{h}$  are crucial. From a Bayesian point of view, those restrictions are modeled as prior distributions on the latent clean image and the blur.

This work was supported by grant PID2022-140189OB-C22 funded by MCIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe”, by the European Union. The work by Francisco M. Castro-Macías was supported by MCIN under FPU contract FPU21/01874. The work by Fernando Pérez-Bueno was supported in part by JDC2022-048784-I, funded by MCIN/AEI/10.13039/501100011033 and the European Union “NextGenerationEU”/PRTR.

A common approach is to build upon the well-known fact that, when high-pass filters are applied to natural images, the resulting coefficients are sparse, which means that most of them are close to zero while only a few are large [5]. State of the art BID methods exploit this property using sparsity promoting image priors such as Super-Gaussian (SG) priors [6]. These distributions have heavier tails than the Gaussian, are more peaked, and have a positive excess of kurtosis. See [7] for a complete review on SG distributions. Although they have proven to be a valuable tool to introduce prior information in BID, some of the most widely SG priors employed are not differentiable around zero, which hampers the optimization process. Previous works have attempted to modify their behavior around zero to make them numerically stable [8]. To account for this problem, and for the first time in the BID literature, we propose to use the Hyperbolic Secant (HS) distribution [9]. Among its appealing theoretical properties [10], the HS distribution is well-behaved around zero, which helps stabilize the variational optimization process. It admits a Gaussian Scale Mixture (GSM) representation [11], [12], which implies that it is SG [13]. Moreover, this representation allows us to augment the joint model with additional variables, that are later marginalized to recover the original model. Using this technique, we develop an estimation procedure for both the blur kernel and the image using variational Bayesian inference. Finally, the proposed method is compared against classical and state of the art BID techniques.

The rest of the paper is organized as follows: the BID state of the art methods and the HS distribution applications are briefly reviewed in Section II. In Section III, we model the BID problem using the Bayesian framework, introducing the HS distribution as a prior. Then, in Section IV, we use variational Bayes and the GSM representation of the HS distribution to propose an estimation method for the unknown kernel and the original image. The performance of the proposed method is evaluated in Section V. Finally, Section VI concludes the paper.

## II. RELATED WORK

Traditionally, BID algorithms have employed an analytical strategy that explicitly describes the forward model, selects the criteria for getting a solution, and decides on an optimization technique. Regularization based techniques typically try to optimize a criterion, such as minimizing the  $\ell_2$  error norm  $\|\mathbf{y} - \mathbf{Hx}\|^2$ , subject to some restrictions based on prior (or domain) knowledge that is incorporated into the solution process. Stochastic techniques treat the unknowns (blur, clean image, and parameters) as stochastic quantities, using maximum likelihood, maximum *a posteriori* (MAP) or fully (hierarchical) Bayesian approaches to estimate them. See [14] for a review on Bayesian BID.

A huge effort has been devoted to designing priors that constrain the solutions to those that are close to the real ones. For the blur, flat priors have been usually used, along with a Dirichlet prior [15] and a Simultaneous Autoregressive (SAR) prior [16]. For the clean latent image, the first works used smooth priors such as Conditional Autoregressive, SAR and Total Variation [17], [18]. These impose smoothness on the solutions, controlling noise amplification, but also tend to smooth out the details of the image. Because of this, the focus shifted towards sparse priors that adapt to the statistics in the filtered space. When natural images are filtered using high-pass filters, most of the coefficients become zero or very small while only a small number of them are large (e.g., at the edges). This behaviour cannot be modeled using the Gaussian distribution. As a consequence, other priors have been proposed: mixture of Gaussians [19], [20], hyper-Laplacian priors [21], adaptive  $\ell_p$  gradient priors [22], log priors [23], log-TV priors [24], Super Gaussian (SG) priors [5], [8], generalized  $\ell_p/\ell_q$  norm-based priors [25], and approximations of the  $\ell_0$  function [26]. These priors have a drawback: most of them are not differentiable around zero, so the optimization process becomes difficult. Few attempts have been proposed to address this issue, see Huber SG [8]. Thus, the use of other SG distributions that are differentiable around zero, such as the Hyperbolic Secant (HS) distribution proposed in this paper, are of interest since they ease the inference process.

In recent years, models based on Deep Neural Networks (DNNs) have become a popular alternative to solve inverse problems [27]. DNNs use a large number of images to learn a mapping function from the degraded image to the restored image and, possibly, the blur. This makes them faster than approaches requiring iterative optimization for each new sample. However, they lack the flexibility of analytical models, which significantly hampers their effectiveness in BID due to the high variation in the degradations caused by the combination of all the possible blurs and noise [28]. A major effort is being made by the DNN community to tackle the BID problem. See [29] and [30] for a recent survey of deep image deblurring. In [31] a deep Convolutional Neural Network (CNN) was used to learn a regularizer that discriminates sharp from blurred images. In [32] the authors propose a multi-scale residual CNN to eliminate non-uniform motion blur from images. This

model is further refined in [33] where learnable parameters are shared, and in [34] to handle multi-scale blur with low computational complexity. Following a similar approach, [35] proposes a multi-stage progressive image restoration network to recover blurred images. In [36] the unconstrained neural optimization solution to blind deconvolution (SelfDeblur) was proposed. The use of a Generative Adversarial Network (GAN) is proposed in [37] and enhanced in [38]. The method proposed in [39] is built upon these approaches and introduces a scale recurrent conditional GAN. Cycle GANs have also been proposed in [40] and [41] as an approximation to learn how to deblur images by learning how to produce realistic motion blur. In [42], three parametric blur models are estimated using a General Regression Neural Network together with an additional CNN, which is used to select one of the three models for a given observed image.

In this paper, we will use an analytical approach to BID with a HS distribution as a prior for the filtered image and a flat prior for the blur. A family of generalized HS distributions was first derived in [43]. Later, [44] connected it to Brownian Motion, and [9] discussed moments, cumulants, and statistical inference. In terms of applications, [10] showed two experimental studies where the HS distribution fits the tails better than the Normal distribution, and [45] presented three problems in which it appears. During the last few years, the HS distribution has become popular in the context of financial return data [46]. Note that the present work is the first in adapting the HS distribution to the BID problem.

## III. BAYESIAN MODELING

In this section, we first justify the use of the HS distribution as a prior model for the BID problem. Then, we describe the full Bayesian model.

### A. The Hyperbolic Secant distribution as a prior

We consider the following density,

$$f(x; \alpha) = \frac{\alpha}{\pi} \operatorname{sech}(\alpha x), \quad \forall x \in \mathbb{R}, \quad (2)$$

where  $\alpha > 0$ . Its associated distribution belongs to the family of Generalized Hyperbolic Secant distributions [9]. Also, it belongs to the family of  $z$  distributions presented in [12]. In that same work, it is proved that the density of the  $z$  distributions can be expressed as a GSM. As a consequence,  $f$  is SG and it is suitable to be used as a prior for BID. A SG density  $f$  can be expressed as  $f(x) \propto \exp(-\rho(x))$ , where  $\rho$  is a *penalty* function (symmetric around zero and such that  $s \in (0, +\infty) \mapsto \rho(s^{1/2})$  is increasing and concave). Note that  $\rho(x) = \log(\operatorname{sech}(\alpha x))$  is the penalty function of the HS density. See Figure 1 for a comparison between the HS prior and other priors previously used in the literature [5].

In this work, the GSM representation of  $f$  is of great importance: there exists a *mixing* density  $\hat{f}(\omega; \alpha)$  such that

$$f(x; \alpha) = \int_0^{+\infty} \mathcal{N}(x | 0, \omega^{-1}) \hat{f}(\omega; \alpha) d\omega. \quad (3)$$

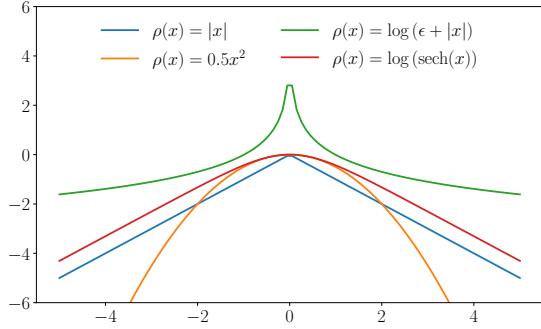


Fig. 1. Penalty functions of the HS density with  $\alpha = 1$  and different SG priors used in the literature [5].

The proof offered in [12], although elegant, does not provide an explicit form for the mixing density. However, it can be obtained using the Pólya-Gamma distribution presented in [47], [48]. See Appendix A for the details.

### B. Modeling the BID problem

We start from the degradation model given by Eq. (1). We formulate the BID problem in the filter space [49], [50], by applying  $N$  high-pass filters  $\{\mathbf{F}_n\}_{n=1}^N$  to the blurred noisy image to obtain  $N$  pseudo-observations

$$\mathbf{y}_n = \mathbf{F}_n \mathbf{y} = \mathbf{H} \mathbf{F}_n \mathbf{x} + \mathbf{F}_n \boldsymbol{\eta} = \mathbf{H} \mathbf{x}_n + \boldsymbol{\eta}_n \quad (4)$$

where  $\mathbf{x}_n = \mathbf{F}_n \mathbf{x}$ ,  $\boldsymbol{\eta}_n = \mathbf{F}_n \boldsymbol{\eta}$ , and  $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \beta^{-1} \mathbf{I})$ . It is assumed that  $\beta > 0$  is known. We denote the set of filtered versions by  $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$  and  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ . Then, from Eq. (4), we obtain,

$$p(\mathbf{y}_n | \mathbf{x}_n, \mathbf{h}, \beta_n) = \mathcal{N}(\mathbf{y}_n | \mathbf{H} \mathbf{x}_n, \beta_n^{-1} \mathbf{I}), \quad (5)$$

where we use the approximation  $\beta_n^{-1} \mathbf{I} \approx \beta^{-1} \mathbf{F}_n \mathbf{F}_n^\top$  and

$$p(\mathbf{Y} | \mathbf{X}, \mathbf{h}, \boldsymbol{\beta}) = \prod_{n=1}^N p(\mathbf{y}_n | \mathbf{x}_n, \mathbf{h}, \beta_n), \quad (6)$$

where  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_N]^\top$ .

We adopt the HS distribution, described in Sec. III-A, as a prior distribution on the pixels of the filtered images, that is,

$$p(\mathbf{X} | \boldsymbol{\alpha}) \propto \prod_{n=1}^N \prod_{i=1}^{HW} \text{sech}(\alpha_n x_n^i), \quad (7)$$

where  $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_N]^\top$  and  $\mathbf{x}_n = [x_n^1, \dots, x_n^{HW}]^\top$ .

Finally, we assume a flat prior  $p(\mathbf{h}) \propto \text{const}$  for the blur. With these ingredients, the joint probability distribution is

$$p(\mathbf{Y}, \mathbf{X}, \mathbf{h} | \boldsymbol{\alpha}, \boldsymbol{\beta}) = p(\mathbf{h}) p(\mathbf{Y} | \mathbf{X}, \mathbf{h}, \boldsymbol{\beta}) p(\mathbf{X} | \boldsymbol{\alpha}). \quad (8)$$

### IV. BAYESIAN INFERENCE

Our approach involves estimating the unknown blur kernel  $\mathbf{h}$  in the filtered image space, followed by estimating the original image using a non-blind deconvolution algorithm. Following the Bayesian perspective, we aim to utilize the posterior distribution  $p(\mathbf{X}, \mathbf{h} | \mathbf{Y}, \boldsymbol{\alpha}, \boldsymbol{\beta})$  to infer  $\mathbf{h}$ . Since it

can not be obtained in closed form, we approximate it using variational Bayesian inference [51]. However, the presence of the HS prior makes it impossible to obtain a closed form expression of the variational posterior. To overcome this issue, we introduce the following augmented model that enables tractable variational inference,

$$p(\mathbf{Y}, \mathbf{X}, \mathbf{h}, \boldsymbol{\omega} | \boldsymbol{\alpha}, \boldsymbol{\beta}) = p(\mathbf{h}) p(\mathbf{Y} | \mathbf{X}, \mathbf{h}, \boldsymbol{\beta}) p(\mathbf{X}, \boldsymbol{\omega} | \boldsymbol{\alpha}), \quad (9)$$

where  $\boldsymbol{\omega} = \{\omega_n^i : n \in \{1, \dots, N\}, i \in \{1, \dots, HW\}\}$  and

$$p(\mathbf{X}, \boldsymbol{\omega} | \boldsymbol{\alpha}) = p(\mathbf{X} | \boldsymbol{\omega}) p(\boldsymbol{\omega} | \boldsymbol{\alpha}), \quad (10)$$

$$p(\mathbf{X} | \boldsymbol{\omega}) \propto \prod_{n=1}^N \prod_{i=1}^{HW} \mathcal{N}(x_n^i | 0, (\omega_n^i)^{-1}), \quad (11)$$

$$p(\boldsymbol{\omega} | \boldsymbol{\alpha}) = \prod_{n=1}^N \prod_{i=1}^{HW} \hat{f}(\omega_n^i; \alpha_n). \quad (12)$$

Integrating in  $\boldsymbol{\omega}$  and using the GSM representation of Eq. (3) we recover the original model. We use the mean-field variational Bayesian approach [51] to approximate  $p(\mathbf{X}, \mathbf{h}, \boldsymbol{\omega} | \mathbf{Y}, \boldsymbol{\alpha}, \boldsymbol{\beta})$  by the variational distribution

$$q(\mathbf{X}, \mathbf{h}, \boldsymbol{\omega}) = q(\mathbf{h}) \prod_{n=1}^N q(\mathbf{x}_n) \left\{ \prod_{i=1}^{HW} q(\omega_n^i) \right\}, \quad (13)$$

by minimizing the Kullback-Leibler (KL) divergence between the variational distribution approximation and the true posterior. The solution for each variational factor is given by [51, Eq. (10.9)]. To estimate the blur kernel we use the mode of  $q(\mathbf{h})$  given by  $\hat{\mathbf{h}} = \text{argmax}_{\mathbf{h}} \log q(\mathbf{h})$ . Let us now specify each estimate.

#### A. Filtered image estimation

From [51, Eq. (10.9)] we obtain

$$\begin{aligned} \log q(\mathbf{x}_n) &= \log p(\mathbf{y}_n | \mathbf{x}_n, \mathbf{h}) + \mathbb{E}_{q(\boldsymbol{\omega})} [p(\mathbf{x} | \boldsymbol{\omega})] + \text{const} \\ &= -\frac{1}{2} \mathbf{x}_n^\top (\beta_n \mathbf{H}^\top \mathbf{H} + \boldsymbol{\Theta}) \mathbf{x}_n + \\ &\quad + \beta_n \mathbf{y}_n^\top \mathbf{H} \mathbf{x}_n + \text{const} \end{aligned} \quad (14)$$

where  $\boldsymbol{\Theta} = \mathbb{E}_{q(\boldsymbol{\omega})} [\text{diag}(\boldsymbol{\omega})]$ . Therefore,

$$q(\mathbf{x}_n) = \mathcal{N}(\mathbf{x}_n | \mathbf{m}_{\mathbf{x}_n}, \boldsymbol{\Sigma}_{\mathbf{x}_n}), \quad (16)$$

$$\boldsymbol{\Sigma}_{\mathbf{x}_n}^{-1} = \beta_n \mathbf{H}^\top \mathbf{H} + \boldsymbol{\Theta}, \quad \mathbf{m}_{\mathbf{x}_n} = \beta_n \boldsymbol{\Sigma}_{\mathbf{x}_n} \mathbf{H}^\top \mathbf{y}_n. \quad (17)$$

To avoid the inversion of  $\boldsymbol{\Sigma}_{\mathbf{x}_n}^{-1}$ , we compute the mean by solving the following linear system using the Conjugate Gradient (CG) method,

$$\boldsymbol{\Sigma}_{\mathbf{x}_n}^{-1} \mathbf{m}_{\mathbf{x}_n} = \beta_n \mathbf{H}^\top \mathbf{y}_n. \quad (18)$$

#### B. Blur estimation

For the blur, applying [51, Eq. (10.9)] we obtain

$$\log q(\mathbf{h}) = \sum_{n=1}^N \mathbb{E}_{q(\mathbf{x}_n)} [\log p(\mathbf{y}_n | \mathbf{x}_n, \mathbf{h})] + \text{const}. \quad (19)$$

We estimate the blur as the mode of  $q(\mathbf{h})$ , which results in

$$\hat{\mathbf{h}} = \arg \min_{\mathbf{h} \in \Delta^K} \{ \mathbf{h}^\top \mathbf{C}_h \mathbf{h} - 2 \mathbf{h}^\top \mathbf{b}_h \}, \quad (20)$$

where  $\Delta^K = \{(h_1, \dots, h_K) \in \mathbb{R}^K : h_i \geq 0, \sum_i h_i = 1\}$ ,  $\mathbf{C}_h \in \mathbb{R}^{K \times K}$  and  $\mathbf{b}_h \in \mathbb{R}^K$  are given by

$$[\mathbf{C}_h]_{ij} = \sum_{n=1}^N \sum_{l=1}^{HW} \left\{ [\mathbf{m}_{x_n}]_{i+l} [\mathbf{m}_{x_n}]_{j+l} + [\Sigma_{x_n}]_{i+l, j+l} \right\}, \quad (21)$$

$$[\mathbf{b}_h]_i = \sum_{n=1}^N \sum_{l=1}^{HW} [\mathbf{m}_{x_n}]_{i+l} [\mathbf{y}_n]_l, \quad (22)$$

where we have used the operator  $[.]_i$  to denote the  $i$ -th component of a vector or a matrix. To avoid the inversion of  $\Sigma_{x_n}^{-1}$ , we apply the Jacobi approximation to invert only its diagonal. As Eq. (20) is a quadratic, it can be solved efficiently.

### C. Estimation of the augmentation variables $\omega$

Observe that the full distribution  $q(\omega_n^i)$  is not needed to estimate  $q(x_n)$  and  $\hat{\mathbf{h}}$ , as only the expectation  $\mathbb{E}_{q(\omega_n^i)} [\omega_n^i]$  is required to calculate  $\Sigma_{x_n}^{-1}$ . To compute this expectation in closed form, we adapt the procedure employed in [5], [13]. First, applying [51, Eq. (10.9)] we obtain

$$\log q(\omega_n^i) = \mathbb{E}_{q(x_n^i)} \left[ \log \mathcal{N} \left( x_n^i | 0, (\omega_n^i)^{-1} \right) \right] + \quad (23)$$

$$+ \log \hat{f}(\omega_n^i; \alpha_n) + \text{const.} \quad (24)$$

Next, we observe that  $\mathbb{E}_{q(x_n^i)} \left[ \log \mathcal{N} \left( x_n^i | 0, (\omega_n^i)^{-1} \right) \right] = \log \mathcal{N} \left( \xi_n^i | 0, (\omega_n^i)^{-1} \right)$ , with  $\xi_n^i = \sqrt{\mathbb{E}_{q(x_n^i)} [(x_n^i)^2]}$ . Therefore, we obtain  $q(\omega_n^i) \propto \mathcal{N} \left( \xi_n^i | 0, (\omega_n^i)^{-1} \right) \hat{f}(\omega_n^i; \alpha_n)$ . Normalizing, we arrive at

$$q(\omega_n^i) f(\xi_n^i; \alpha_n) = \mathcal{N} \left( \xi_n^i | 0, (\omega_n^i)^{-1} \right) \hat{f}(\omega_n^i; \alpha_n). \quad (25)$$

Finally, we differentiate under the integral sign in Eq. (3) and use Eq. (25) to obtain

$$\mathbb{E}_{q(\omega_n^i)} [\omega_n^i] = \frac{\alpha_n \tanh(\alpha_n \xi_n^i)}{\xi_n^i}. \quad (26)$$

### D. Estimation of the original image

Standard mean-field variational inference consists of alternating the estimates of  $\{\mathbf{m}_{x_n}\}_{n=1}^N$ ,  $\{\Sigma_{x_n}^{-1}\}_{n=1}^N$ ,  $\Theta$ , and  $\mathbf{h}$ . This way, we obtain estimations of the filtered images and the blur, but not of the original image. The estimated blur can be used with a non-blind image restoration algorithm to estimate the original image. Following the approach of previous works [8], [52], we compute the original image estimate  $\hat{\mathbf{x}}$  by solving

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{Hx} - \mathbf{y}\|^2 + \lambda \sum_{n=1}^N \|\mathbf{F}_n \mathbf{x}\|^p \right\}, \quad (27)$$

where  $p = 0.8$  [52]. To solve this non-convex optimization problem we use an iteratively reweighted least squares approach (see [52] for the details).

The proposed method is summarized in Algorithm 1. The main loop alternates between the estimates of the filtered images, the hyperparameters, and the blur. To estimate the blur kernel, we apply the pyramid coarse-to-fine approach

suggested in [20]. Once the blur is estimated, the original image is estimated by the non-blind image deconvolution method in Eq. (27).

---

### Algorithm 1: Blind Deconvolution using the Hyperbolic-Secant prior.

---

**Input:** Degraded image  $\mathbf{y}$ , noise parameter  $\beta$ , filters  $\{\mathbf{F}_n\}_{n=1}^N$ , prior parameter  $\alpha$ , number of iterations  $T$ .

- 1 Initialize  $\mathbf{m}_{x_n} = \mathbf{y}_n$  and  $\Sigma_{x_n}^{-1} = \mathbf{0}$  for  $n = 1, \dots, N$ .
- 2 **for**  $t = 1, \dots, T$  **do**
- 3     Compute  $\Theta$  using Eq. (26).
- 4     Compute  $\{\Sigma_{x_n}^{-1}\}_{n=1}^N$  using Eq. (17) and  $\{\mathbf{m}_{x_n}\}_{n=1}^N$  by solving the system in Eq. (18).
- 5     Compute  $\hat{\mathbf{h}}$  by solving Eq. (20).
- 6 Compute  $\hat{\mathbf{x}}$  by solving Eq. (27).

**Output:** Original image estimation  $\hat{\mathbf{x}}$ .

---

## V. EXPERIMENTAL RESULTS

In this section, we assess the quality of the proposed approach and compare its results with those obtained using the following methods: MoG [20], SG in [5] and [53], Huber SG [8], ECP [26], Li [31] and SelfDeblur [36]. For the SG priors in [5], we consider  $\log$ ,  $\ell_1$  and  $\exp$ .

We use the Levin dataset [50], which is widely utilized as a workbench for testing many BID algorithms. The different methods are compared quantitatively and qualitatively. As quantitative metrics, we use the Sum of Square Distances (SSD) between the deconvolved and original images, the Peak Signal to Noise Ratio (PSNR), and the Structural Similarity Index (SSIM) [54]. If the BID method allows to obtain non-blind deconvolutions, we present both blind measures (SSDb, PSNRb, SSIMb) and non-blind measures (SSDnb, PSNRnb, SSIMnb). These non-blind measures are obtained using the ground truth kernel. Furthermore, we show the ratio SSDb/SSDnb. For PSNR and SSIM higher values correspond to better results, while for SSD and ratio, the lower the better.

The filters used in Eq. (4) correspond to the first order horizontal and vertical difference filters. For all images in the dataset, we used  $\beta = 10^4$ ,  $\alpha_1 = 10^{2.4}$ , and  $\alpha_2 = 10^{2.15}$ . Those parameters were selected by grid search to obtain the best results in one image and, then, used for all the experiments. All competing methods were run using the parameters suggested by the authors for this dataset.

Tables I and II show the mean quality metrics obtained for the whole dataset using the different methods. We also include the mean execution time per image on an Intel Xeon Silver 4314 2.40 GHz CPU for the different analytical methods and on a Nvidia GeForce RTX 3090 GPU for DNN methods. The best results are highlighted in bold. The proposed HS method obtains better results than the rest of the analytical methods (Table I), except for the execution time which is lower for the Huber SG method and the ratio SSDb/SSDnb for the ECP method. This is due to the poorer performance of the ECP method for non-blind deconvolution. When we compare with DNN methods (Table II), SelfDeblur shows the best SSD and PSNR values while the proposed HS achieves the best SSIM

TABLE I  
COMPARISON BETWEEN HS AND OTHER ANALYTICAL METHODS FOR LEVIN DATASET.  
( $\downarrow$ ) MEANS LOWER IS BETTER, ( $\uparrow$ ) MEANS HIGHER IS BETTER.

| Method        | SSDb ( $\downarrow$ ) | PSNRb ( $\uparrow$ ) | SSIMb ( $\uparrow$ ) | SSDnb ( $\downarrow$ ) | PSNRnb ( $\uparrow$ ) | SSIMnb ( $\uparrow$ ) | ratio ( $\downarrow$ ) | CPU (secs.) ( $\downarrow$ ) |
|---------------|-----------------------|----------------------|----------------------|------------------------|-----------------------|-----------------------|------------------------|------------------------------|
| log [53]      | 152.4                 | 27.9                 | 0.8162               | 41.49                  | 32.1                  | 0.9253                | 4.12                   | 69.4                         |
| $\ell_1$ [53] | 239.1                 | 25.4                 | 0.7391               | 39.17                  | 32.3                  | 0.9211                | 5.93                   | 69.72                        |
| MoG [20]      | 106.1                 | 28.6                 | 0.8428               | 39.53                  | 32.2                  | 0.9237                | 2.79                   | 69.96                        |
| exp [53]      | 101.1                 | 28.6                 | 0.8528               | 43.71                  | 32                    | 0.9234                | 2.59                   | 69.75                        |
| Huber SG [8]  | 110.6                 | 29.9                 | 0.8868               | <b>36.47</b>           | <b>32.8</b>           | <b>0.9406</b>         | 2.67                   | <b>7.418</b>                 |
| ECP [26]      | 70.53                 | 30                   | 0.8957               | 55.98                  | 30.9                  | 0.91                  | <b>1.3</b>             | 100                          |
| HS            | <b>69.01</b>          | <b>30.5</b>          | <b>0.9069</b>        | <b>36.47</b>           | <b>32.8</b>           | <b>0.9406</b>         | 2                      | 68.69                        |

TABLE II

COMPARISON BETWEEN HS AND DNN METHODS FOR LEVIN DATASET.  
( $\downarrow$ ) MEANS LOWER IS BETTER, ( $\uparrow$ ) MEANS HIGHER IS BETTER.

| Method          | SSDb ( $\downarrow$ ) | PSNRb ( $\uparrow$ ) | SSIMb ( $\uparrow$ ) | Time (secs.) ( $\downarrow$ ) |
|-----------------|-----------------------|----------------------|----------------------|-------------------------------|
| Li [31]         | 103.3                 | 29.7                 | 0.8815               | 149 (GPU)                     |
| SelfDeblur [36] | <b>57.8</b>           | <b>31.5</b>          | 0.9029               | 399 (GPU)                     |
| HS              | 69.01                 | 30.5                 | <b>0.9069</b>        | <b>68.69 (CPU)</b>            |

value. The HS method, which is executed in CPU, requires a significantly lower execution time than the two DNN methods, that require a GPU.

Figure 2 shows cumulative SSD histograms of the blind and non-blind deconvolved images and their ratios. The best histograms of SSDb and ratio values are those of HS. In the SSDnb histogram, the proposed HS and Huber SG methods share the best behavior. Note that both methods use the same non-blind image deconvolution method. These results indicate that the HS method provides more accurate estimations of the unknown blur than the other methods in this dataset.

Figure 3 shows a visual comparison of an original image, the degraded image, and the restorations obtained using different methods. In Figure 3(b)-(k), the true blur kernel and the estimated kernels are included in the degraded and the restored images. Visually, the best reconstructions are obtained by the proposed method. Huber SG, Li, and ECP also achieve visually appealing reconstruction, but lack some fine details.

Finally, Figure 4 shows a real degraded color image and the restorations obtained by the best performing methods. For this image we have used as noise variance  $\beta = 0.4 \times 10^5$  and  $\alpha_1 = \alpha_2 = 10^2$ . While all restorations have an acceptable visual quality, those obtained using the HS and Li methods achieve the best results, see Figure 4(f) and (d). In Figure 4(e), SelfDeblur recovers a noisy image, while in Figure 4(b) Huber SG over-smooths the image.

## VI. CONCLUSIONS

In this work, we propose a BID method that uses the HS distribution for the first time in the literature. This seldom explored distribution belongs to the family of SG distributions, has heavy tails and behaves suitably around zero, which makes it a good candidate as a prior distribution in BID. We derive a GSM representation, which is then used to formulate a new BID variational Bayesian method. The proposed method shows competitive or superior performance in the Levin dataset, which supports our claims.

## APPENDIX A THE MIXING DENSITY OF EQ. (3)

For completeness, we derive an explicit form for the mixing density  $f(\omega; \alpha)$  of Eq. (3). First, we note that  $\operatorname{sech}(y) = 2 \exp(y) / (\exp(2y) + 1)$ . Then, from [47, Eq. 2] we obtain, after a change of variable,

$$f(x; \alpha) = \int_0^{+\infty} \mathcal{N}(x | 0, \omega^{-1}) \frac{\operatorname{PG}(\omega / (4\alpha^2) | 0, 1)}{2\alpha\sqrt{2\pi}\omega} d\omega, \quad (28)$$

where  $\operatorname{PG}(\omega | 0, 1)$  denotes the Pólya-Gamma density defined in [47]. Finally, we check that  $\hat{f}(\omega; \alpha) = (2\alpha\sqrt{2\pi}\omega)^{-1} \operatorname{PG}(\omega / (4\alpha^2) | 0, 1)$  is a density. To do so, first observe that  $\hat{f}(\omega; \alpha) \geq 0$ . To show that it integrates to 1, integrate with respect to  $x$  in Eq. (28) and apply the Fubini-Tonelli Theorem [55].

## REFERENCES

- [1] T. E. Bishop, S. D. Babacan, B. Amizic, A. K. Katsaggelos, T. Chan, and R. Molina, “Blind image deconvolution: problem formulation and existing approaches,” in *Blind Image Deconvolution*. CRC press, 2017, pp. 21–62.
- [2] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, “Removing camera shake from a single photograph,” in *ACM Siggraph 2006 Papers*, 2006, pp. 787–794.
- [3] J. Jia, “Single image motion deblurring using transparency,” in *2007 IEEE Conference on computer vision and pattern recognition*, 2007, pp. 1–8.
- [4] C. Zhou, S. Lin, and S. K. Nayar, “Coded aperture pairs for depth from defocus and defocus deblurring.” *International journal of computer vision*, vol. 93, no. 1, 2011.
- [5] S. Babacan, R. Molina, M. Do, and A. Katsaggelos, “Bayesian blind deconvolution with general sparse image priors,” in *European Conference on Computer Vision (ECCV)*, 2012, pp. 341–355.
- [6] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive algorithms and stochastic approximations*. Springer Science & Business Media, 2012, vol. 22.
- [7] J. A. Palmer, K. Kreutz-Delgado, and S. Makeig, “Strong sub-and super-gaussianity,” in *9th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*. Springer, 2010, pp. 303–310.
- [8] X. Zhou, M. Vega, F. Zhou, R. Molina, and A. K. Katsaggelos, “Fast Bayesian blind deconvolution with Huber super Gaussian priors,” *Digital Signal Processing*, vol. 60, pp. 122–133, 2017.
- [9] W. L. Harkness and M. L. Harkness, “Generalized hyperbolic secant distributions,” *Journal of the American Statistical Association*, vol. 63, no. 321, pp. 329–337, 1968, taylor & Francis.
- [10] D. C. Vaughan, “The generalized secant hyperbolic distribution and its properties,” *Communications in Statistics-Theory and Methods*, vol. 31, no. 2, pp. 219–238, 2002.
- [11] D. F. Andrews and C. L. Mallows, “Scale mixtures of normal distributions,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no. 1, pp. 99–102, 1974.

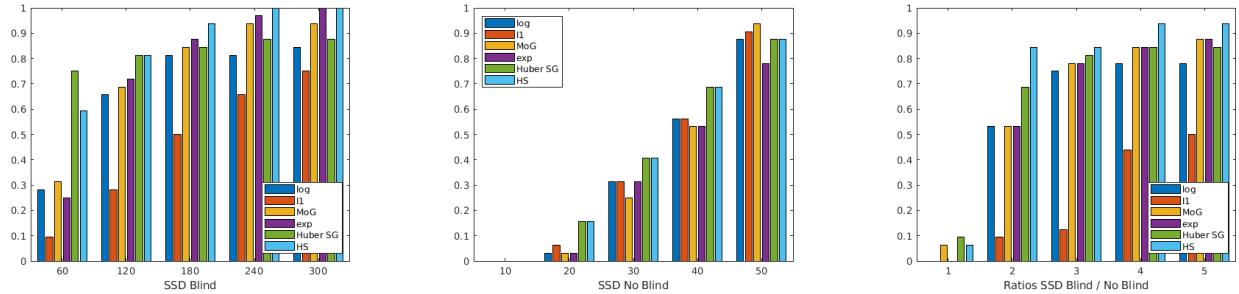


Fig. 2. Cumulative histograms of SSDs on Levin dataset for blind and non-blind deconvolution, and their ratios (right).



Fig. 3. Visual comparison of the results of the different methods for an image of the Levin dataset.

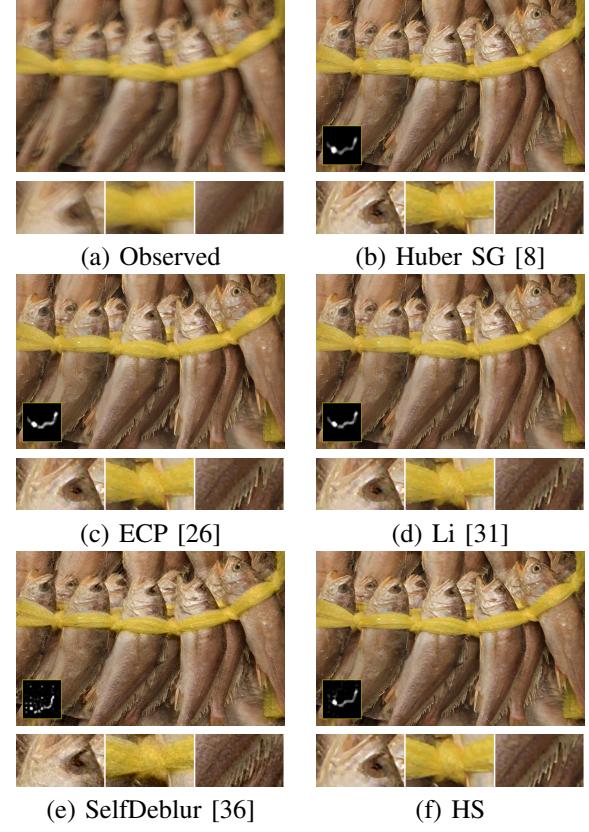


Fig. 4. Blind Deconvolution results on a real image with estimated kernels shown in insets. Zooms of three image areas are shown.

- [12] O. Barndorff-Nielsen, J. Kent, and M. Sørensen, “Normal variance-mean mixtures and z distributions,” *International Statistical Review/Revue Internationale de Statistique*, pp. 145–159, 1982.
- [13] J. A. Palmer, “Variational and scale mixture representations of non-Gaussian densities for estimation in the Bayesian linear model: Sparse coding, independent component analysis, and minimum entropy segmentation,” Ph.D. dissertation, University of California, San Diego, 2006.
- [14] P. Ruiz, X. Zhou, J. Mateos, R. Molina, and A. K. Katsaggelos, “Variational Bayesian blind image deconvolution: A review,” *Digital Signal Processing*, vol. 47, pp. 116–127, 2015.
- [15] X. Zhou, J. Mateos, F. Zhou, R. Molina, and A. K. Katsaggelos, “Variational Dirichlet blur kernel estimation,” *IEEE transactions on Image Processing*, vol. 24, pp. 5127–5139, 2015.
- [16] Y. Huang, E. Chouzenoux, and J.-C. Pesquet, “Unrolled Variational Bayesian algorithm for image blind deconvolution,” *IEEE Transactions on Image Processing*, vol. 32, pp. 430–445, 2023.
- [17] R. Molina, A. K. Katsaggelos, J. Abad, and J. Mateos, “A Bayesian

- approach to blind deconvolution based on Dirichlet distributions,” in *ICASSP*, 1997.
- [18] D. Perrone and P. Favaro, “Total variation blind deconvolution: The devil is in the details,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
- [19] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, “Removing camera shake from a single photograph,” *ACM Trans. Graph*, vol. 25, no. 3, 2006.
- [20] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, “Efficient marginal likelihood optimization in blind deconvolution,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 2657–2664.
- [21] D. Krishnan and R. Fergus, “Fast image deconvolution using hyper-laplacian priors,” in *Conference on Neural Information Processing Systems*, 2009.
- [22] B. Zhang, R. Liu, H. Li, Q. Yuan, X. Fan, and Z. Luo, “Blind image deblurring using adaptive priors,” in *Internet Multimedia Computing and Service*, 2018, pp. 13–22.

- [23] D. Perrone and P. Favaro, "A logarithmic image prior for blind deconvolution," *International Journal of Computer Vision*, vol. 117, no. 2, pp. 159–172, 2016.
- [24] D. Perrone, R. Diethelm, and P. Favaro, "Blind deconvolution via lower-bounded logarithmic image priors," in *International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, 2015.
- [25] L. Chen, Q. Sun, and F. Wang, "Adaptive blind deconvolution using generalized cross-validation with generalized lp/lq norm regularization," *Neurocomputing*, vol. 399, pp. 75–85, 2020.
- [26] Y. Yan, W. Ren, Y. Guo, R. Wang, and X. Cao, "Image deblurring via extreme channels prior," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4003–4011.
- [27] A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos, "Using deep neural networks for inverse problems in imaging," *Signal Processing Magazine*, vol. 35, no. 1, pp. 20–36, 2018.
- [28] S. López-Tapia, J. Mateos, R. Molina, and A. Katsaggelos, "Learning moore-penrose based residuals for robust non-blind image deconvolution," *Digital Signal Processing*, vol. 142, p. 104193, October 2023.
- [29] J. Koh, J. Lee, and S. Yoon, "Single-image deblurring with neural networks: A comparative survey," *Computer Vision and Image Understanding*, vol. 203, p. 103134, 2021.
- [30] K. Zhang, W. Ren, W. Luo, W.-S. Lai, B. Stenger, M.-H. Yang, and H. Li, "Deep image deblurring: A survey," 2022.
- [31] L. Li, J. Pan, W.-S. Lai, C. Gao, N. Sang, and M.-H. Yang, "Blind image deblurring via deep discriminative priors," *International Journal of Computer Vision*, vol. 127, no. 8, pp. 1025–1043, 2019.
- [32] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 257–265.
- [33] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [34] S.-J. Cho, S.-W. Ji, J.-P. Hong, S.-W. Jung, and S.-J. Ko, "Rethinking coarse-to-fine approach in single image deblurring," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2021, pp. 4641–4650.
- [35] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [36] D. Ren, K. Zhang, Q. Wang, Q. Hu, and W. Zuo, "Neural blind deconvolution using deep priors," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3341–3350.
- [37] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8183–8192.
- [38] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 8877–8886.
- [39] J. Liu, W. Sun, and M. Li, "Recurrent conditional generative adversarial network for image deblurring," *IEEE Access*, vol. 7, pp. 6186–6193, 2019.
- [40] K. Zhang, W. Luo, Y. Zhong, L. Ma, B. Stenger, W. Liu, and H. Li, "Deblurring by realistic blurring," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 2737–2746.
- [41] Y. Wen, J. Chen, B. Sheng, Z. Chen, P. Li, P. Tan, and T.-Y. Lee, "Structure-aware motion deblurring using multi-adversarial optimized CycleGAN," *IEEE Transactions on Image Processing*, vol. 30, pp. 6142–6155, 2021.
- [42] R. Yan and L. Shao, "Blind image blur estimation via deep learning," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1910–1921, 2016.
- [43] W. Perks, "On some experiments in the graduation of mortality statistics," *Journal of the Institute of Actuaries*, vol. 63, no. 1, pp. 12–57, 1932.
- [44] J. Talacko, "Perks' distributions and their role in the theory of wiener's stochastic variables," *Trabajos de estadística*, vol. 7, no. 2, pp. 159–174, 1956.
- [45] P. Ding, "Three occurrences of the hyperbolic-secant distribution," *The American Statistician*, vol. 68, no. 1, pp. 32–35, 2014.
- [46] M. J. Fischer, *Generalized hyperbolic secant distributions: with applications to finance*. Springer Science & Business Media, 2013.
- [47] N. G. Polson, J. G. Scott, and J. Windle, "Bayesian inference for logistic models using pólya-gamma latent variables," *Journal of the American Statistical Association*, vol. 108, no. 504, pp. 1339–1349, 2013.
- [48] A. Klami, "Pólya-Gamma augmentations for factor models," in *Asian Conference on Machine Learning*. PMLR, 2015, pp. 112–128.
- [49] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM transactions on graphics*, vol. 27, no. 3, pp. 1–10, 2008.
- [50] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 1964–1971.
- [51] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.
- [52] X. Zhou, F. Zhou, X. Bai, and B. Xue, "A boundary condition based deconvolution framework for image deblurring," *Journal of Computational and Applied Mathematics*, vol. 261, pp. 14–29, 2014.
- [53] M. Vega, R. Molina, and A. K. Katsaggelos, "Parameter estimation in Bayesian blind deconvolution with super Gaussian image priors," in *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO)*, Sep. 2014, pp. 1632–1636.
- [54] 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.
- [55] H. L. Royden and P. Fitzpatrick, *Real analysis*. Macmillan New York, 1988, vol. 32.