



UNIVERSIDAD
DE GRANADA



NORTHWESTERN
UNIVERSITY

Bayesian Blind Image Deconvolution using a Hyperbolic-Secant prior

F.M. Castro-Macías, F. Pérez-Bueno, M. Vega, J. Mateos, R. Molina, A. K.
Katsaggelos

2024 IEEE International Conference on Image Processing

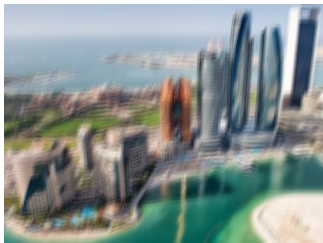
Overview

1. Blind Image Deconvolution
2. The Hyperbolic Secant (HS) distribution
3. Modelling and inference
4. Results
5. Conclusions

Plan

1. Blind Image Deconvolution
2. The Hyperbolic Secant (HS) distribution
3. Modelling and inference
4. Results
5. Conclusions

Blind Image Deconvolution (BID)



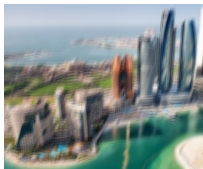
y



x

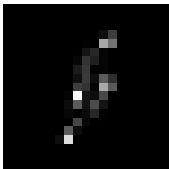
Blind Image Deconvolution (BID)

$$\mathbf{y} = \mathbf{h} * \mathbf{x} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \beta^{-1} \mathbf{I})$$



\mathbf{y}

=



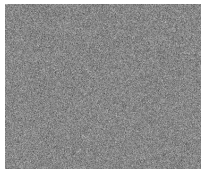
\mathbf{h}

*



\mathbf{x}

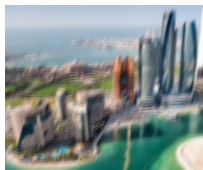
+



$\boldsymbol{\eta}$

Blind Image Deconvolution (BID)

$$\mathbf{y} = \mathbf{h} * \mathbf{x} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \beta^{-1} \mathbf{I})$$



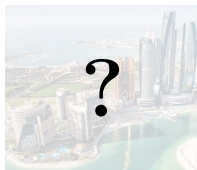
\mathbf{y}

=



\mathbf{h}

*



\mathbf{x}

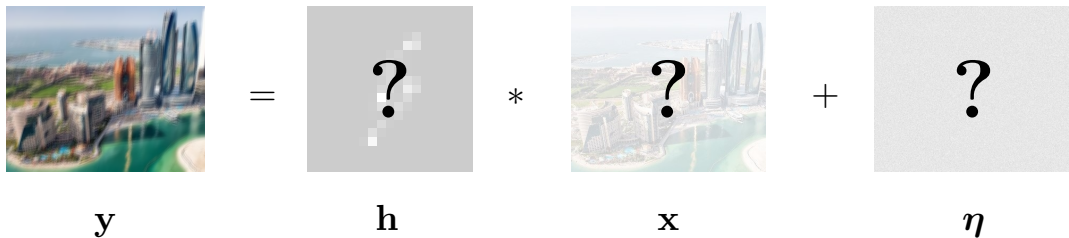
+



$\boldsymbol{\eta}$

Blind Image Deconvolution (BID)

$$\mathbf{y} = \mathbf{h} * \mathbf{x} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \beta^{-1} \mathbf{I})$$



BID is ill-conditioned

Given \mathbf{y} , there are infinitely many (\mathbf{x}, \mathbf{h}) such that the above equation holds... we need to restrict the solution space.

Idea: sparsity

When a high pass filter is applied to a sharp image, the resulting image is **sparse**.

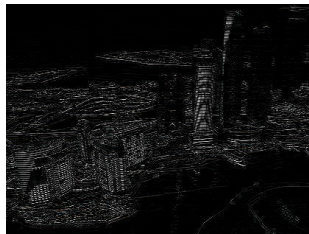
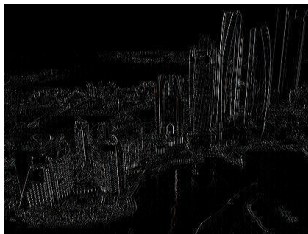


Figure: Clean image and the resulting filtered images.

We need to look for solutions with this property! How? Using Super Gaussian (SG) priors!

Plan

1. Blind Image Deconvolution
- 2. The Hyperbolic Secant (HS) distribution**
3. Modelling and inference
4. Results
5. Conclusions

The Hyperbolic Secant (HS) distribution

The Hyperbolic Secant (HS) density is given by

$$f(x; \alpha) = \frac{\alpha}{\pi} \operatorname{sech}(\alpha x) = \frac{2\alpha}{\pi} (e^{\alpha x} + e^{-\alpha x})^{-1}, \quad \forall x \in \mathbb{R}.$$

The Hyperbolic Secant (HS) distribution

The Hyperbolic Secant (HS) density is given by

$$f(x; \alpha) = \frac{\alpha}{\pi} \operatorname{sech}(\alpha x) = \frac{2\alpha}{\pi} (e^{\alpha x} + e^{-\alpha x})^{-1}, \quad \forall x \in \mathbb{R}.$$

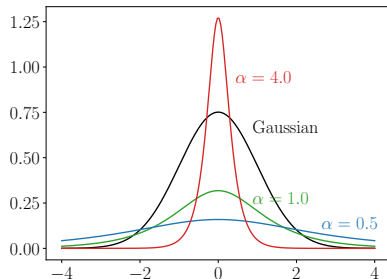
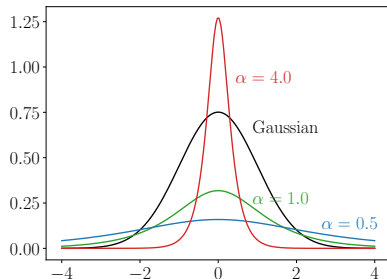


Figure: Gaussian and HS distributions.

The Hyperbolic Secant (HS) distribution

The Hyperbolic Secant (HS) density is given by

$$f(x; \alpha) = \frac{\alpha}{\pi} \operatorname{sech}(\alpha x) = \frac{2\alpha}{\pi} (e^{\alpha x} + e^{-\alpha x})^{-1}, \quad \forall x \in \mathbb{R}.$$



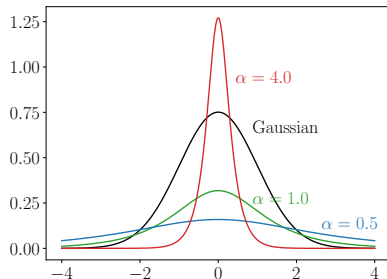
- Appears in different contexts: financial mathematics, pricing of options, and binary classification.

Figure: Gaussian and HS distributions.

The Hyperbolic Secant (HS) distribution

The Hyperbolic Secant (HS) density is given by

$$f(x; \alpha) = \frac{\alpha}{\pi} \operatorname{sech}(\alpha x) = \frac{2\alpha}{\pi} (e^{\alpha x} + e^{-\alpha x})^{-1}, \quad \forall x \in \mathbb{R}.$$



- Appears in different contexts: financial mathematics, pricing of options, and binary classification.
- Never used in BID!

Figure: Gaussian and HS distributions.

The Hyperbolic Secant (HS) distribution

The Hyperbolic Secant (HS) density is given by

$$f(x; \alpha) = \frac{\alpha}{\pi} \operatorname{sech}(\alpha x) = \frac{2\alpha}{\pi} (e^{\alpha x} + e^{-\alpha x})^{-1}, \quad \forall x \in \mathbb{R}.$$

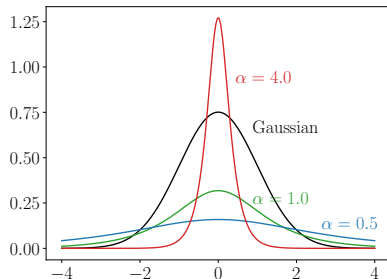


Figure: Gaussian and HS distributions.

- Appears in different contexts: financial mathematics, pricing of options, and binary classification.
- Never used in BID!
- Connection to Brownian motion and the Pólya-Gamma distribution.

The Hyperbolic Secant (HS) distribution

The Hyperbolic Secant (HS) density is given by

$$f(x; \alpha) = \frac{\alpha}{\pi} \operatorname{sech}(\alpha x) = \frac{2\alpha}{\pi} (e^{\alpha x} + e^{-\alpha x})^{-1}, \quad \forall x \in \mathbb{R}.$$

Two important properties

1. **Gaussian Scale Mixture (GSM)** representation: there exists a *mixing* density $\hat{f}(\omega; \alpha)$ such that

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

As a consequence, f is a **Super Gaussian!**

2. f is differentiable around zero (contrary to previously used SGs!)

Plan

1. Blind Image Deconvolution
2. The Hyperbolic Secant (HS) distribution
- 3. Modelling and inference**
4. Results
5. Conclusions

Modelling the problem

1. Consider a set of high-pass filters $\{\mathbf{F}_n\}_{n=1}^N$ and apply them to obtain a set of *pseudo-observations*,

$$\underbrace{\mathbf{F}_n \mathbf{y}}_{\mathbf{y}_n} = \mathbf{h} * \underbrace{\mathbf{F}_n \mathbf{x}}_{\mathbf{x}_n} + \underbrace{\mathbf{F}_n \boldsymbol{\eta}}_{\boldsymbol{\eta}_n}.$$

Modelling the problem

1. Consider a set of high-pass filters $\{\mathbf{F}_n\}_{n=1}^N$ and apply them to obtain a set of *pseudo-observations*,

$$\underbrace{\mathbf{F}_n \mathbf{y}}_{\mathbf{y}_n} = \mathbf{h} * \underbrace{\mathbf{F}_n \mathbf{x}}_{\mathbf{x}_n} + \underbrace{\mathbf{F}_n \boldsymbol{\eta}}_{\boldsymbol{\eta}_n}.$$

2. The components of the probabilistic model are

Modelling the problem

1. Consider a set of high-pass filters $\{\mathbf{F}_n\}_{n=1}^N$ and apply them to obtain a set of *pseudo-observations*,

$$\underbrace{\mathbf{F}_n \mathbf{y}}_{\mathbf{y}_n} = \mathbf{h} * \underbrace{\mathbf{F}_n \mathbf{x}}_{\mathbf{x}_n} + \underbrace{\mathbf{F}_n \boldsymbol{\eta}}_{\boldsymbol{\eta}_n}.$$

2. The components of the probabilistic model are

$$p(\{\mathbf{y}_n\} \mid \{\mathbf{x}_n\}, \mathbf{h}, \boldsymbol{\beta}) = \prod_{n=1}^N \mathcal{N}(\mathbf{y}_n \mid \mathbf{h} * \mathbf{x}_n, \beta_n^{-1} \mathbf{I}),$$

Modelling the problem

1. Consider a set of high-pass filters $\{\mathbf{F}_n\}_{n=1}^N$ and apply them to obtain a set of *pseudo-observations*,

$$\underbrace{\mathbf{F}_n \mathbf{y}}_{\mathbf{y}_n} = \mathbf{h} * \underbrace{\mathbf{F}_n \mathbf{x}}_{\mathbf{x}_n} + \underbrace{\mathbf{F}_n \boldsymbol{\eta}}_{\boldsymbol{\eta}_n}.$$

2. The components of the probabilistic model are

$$p(\{\mathbf{y}_n\} \mid \{\mathbf{x}_n\}, \mathbf{h}, \boldsymbol{\beta}) = \prod_{n=1}^N \mathcal{N}(\mathbf{y}_n \mid \mathbf{h} * \mathbf{x}_n, \beta_n^{-1} \mathbf{I}),$$

$$p(\{\mathbf{x}_n\} \mid \boldsymbol{\alpha}) \propto \prod_{n=1}^N \prod_{i=1}^{HW} \text{sech}(\alpha_n x_n^i), \leftarrow \text{Sparsity!}$$

Modelling the problem

1. Consider a set of high-pass filters $\{\mathbf{F}_n\}_{n=1}^N$ and apply them to obtain a set of *pseudo-observations*,

$$\underbrace{\mathbf{F}_n \mathbf{y}}_{\mathbf{y}_n} = \mathbf{h} * \underbrace{\mathbf{F}_n \mathbf{x}}_{\mathbf{x}_n} + \underbrace{\mathbf{F}_n \boldsymbol{\eta}}_{\boldsymbol{\eta}_n}.$$

2. The components of the probabilistic model are

$$p(\{\mathbf{y}_n\} \mid \{\mathbf{x}_n\}, \mathbf{h}, \boldsymbol{\beta}) = \prod_{n=1}^N \mathcal{N}(\mathbf{y}_n \mid \mathbf{h} * \mathbf{x}_n, \beta_n^{-1} \mathbf{I}),$$

$$p(\{\mathbf{x}_n\} \mid \boldsymbol{\alpha}) \propto \prod_{n=1}^N \prod_{i=1}^{HW} \text{sech}(\alpha_n x_n^i), \leftarrow \text{Sparsity!}$$
$$p(\mathbf{h}) \propto \text{const}.$$

Modelling the problem

1. Consider a set of high-pass filters $\{\mathbf{F}_n\}_{n=1}^N$ and apply them to obtain a set of *pseudo-observations*,

$$\underbrace{\mathbf{F}_n \mathbf{y}}_{\mathbf{y}_n} = \mathbf{h} * \underbrace{\mathbf{F}_n \mathbf{x}}_{\mathbf{x}_n} + \underbrace{\mathbf{F}_n \boldsymbol{\eta}}_{\boldsymbol{\eta}_n}.$$

2. The components of the probabilistic model are

$$p(\{\mathbf{y}_n\} \mid \{\mathbf{x}_n\}, \mathbf{h}, \boldsymbol{\beta}) = \prod_{n=1}^N \mathcal{N}(\mathbf{y}_n \mid \mathbf{h} * \mathbf{x}_n, \beta_n^{-1} \mathbf{I}),$$

$$p(\{\mathbf{x}_n\} \mid \boldsymbol{\alpha}) \propto \prod_{n=1}^N \prod_{i=1}^{HW} \text{sech}(\alpha_n x_n^i), \leftarrow \text{Sparsity!}$$
$$p(\mathbf{h}) \propto \text{const}.$$

3. The joint probabilistic model is given by

$$p(\{\mathbf{y}_n\}, \{\mathbf{x}_n\}, \mathbf{h} \mid \boldsymbol{\beta}, \boldsymbol{\alpha}) = p(\{\mathbf{y}_n\} \mid \{\mathbf{x}_n\}, \mathbf{h}, \boldsymbol{\beta}) p(\{\mathbf{x}_n\} \mid \boldsymbol{\alpha}) p(\mathbf{h})$$

Inference

We aim to use mean-field variational inference to approximate

$$p(\{\mathbf{x}_n\}, \mathbf{h} \mid \{\mathbf{y}_n\}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \approx q(\{\mathbf{x}_n\}) q(\mathbf{h}).$$

Inference

We aim to use mean-field variational inference to approximate

$$p(\{\mathbf{x}_n\}, \mathbf{h} \mid \{\mathbf{y}_n\}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \approx q(\{\mathbf{x}_n\}) q(\mathbf{h}).$$

Problem: the closed form updates involve $\mathbb{E}_{q(\{\mathbf{x}_n\})} [\log \text{sech}(\alpha_n x_n^i)] \rightarrow \text{Intractable!}$

Inference

We aim to use mean-field variational inference to approximate

$$p(\{\mathbf{x}_n\}, \mathbf{h} \mid \{\mathbf{y}_n\}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \approx q(\{\mathbf{x}_n\}) q(\mathbf{h}).$$

Problem: the closed form updates involve $\mathbb{E}_{q(\{\mathbf{x}_n\})} [\log \text{sech}(\alpha_n x_n^i)] \rightarrow \text{Intractable!}$

Solution: Augmented prior on the filtered images

$$p\left(\{\mathbf{x}_n\}_{n=1}^N \mid \boldsymbol{\omega}\right) \propto \prod_{n=1}^N \prod_{i=1}^{HW} \mathcal{N}\left(x_n^i \mid 0, (\omega_n^i)^{-1}\right), \quad p(\boldsymbol{\omega} \mid \boldsymbol{\alpha}) = \prod_{n=1}^N \prod_{i=1}^{HW} \hat{f}(\omega_n^i; \alpha_n)$$

Inference

We aim to use mean-field variational inference to approximate

$$p(\{\mathbf{x}_n\}, \mathbf{h} \mid \{\mathbf{y}_n\}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \approx q(\{\mathbf{x}_n\}) q(\mathbf{h}).$$

Problem: the closed form updates involve $\mathbb{E}_{q(\{\mathbf{x}_n\})} [\log \text{sech}(\alpha_n x_n^i)] \rightarrow \text{Intractable!}$

Solution: Augmented prior on the filtered images

$$p\left(\{\mathbf{x}_n\}_{n=1}^N \mid \boldsymbol{\omega}\right) \propto \prod_{n=1}^N \prod_{i=1}^{HW} \mathcal{N}\left(x_n^i \mid 0, (\omega_n^i)^{-1}\right), \quad p(\boldsymbol{\omega} \mid \boldsymbol{\alpha}) = \prod_{n=1}^N \prod_{i=1}^{HW} \hat{f}(\omega_n^i; \alpha_n)$$

Because of the GSM representation, we recover the original model integrating in $\boldsymbol{\omega}$,

$$p(\{\mathbf{x}_n\} \mid \boldsymbol{\alpha}) = \int p(\{\mathbf{x}_n\} \mid \boldsymbol{\omega}) p(\boldsymbol{\omega} \mid \boldsymbol{\alpha}) d\boldsymbol{\omega}$$

Inference in the augmented model

$$p(\{\mathbf{x}_n\}, \mathbf{h}, \boldsymbol{\omega} \mid \{\mathbf{y}_n\}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \approx q(\{\mathbf{x}_n\}) q(\mathbf{h}) q(\boldsymbol{\omega})$$

Inference in the augmented model

$$p(\{\mathbf{x}_n\}, \mathbf{h}, \boldsymbol{\omega} \mid \{\mathbf{y}_n\}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \approx q(\{\mathbf{x}_n\}) q(\mathbf{h}) q(\boldsymbol{\omega})$$

- The filtered images are given by $q(\mathbf{x}_n) = \mathcal{N}(\mathbf{x}_n \mid \mathbf{m}_{\mathbf{x}_n}, \boldsymbol{\Sigma}_{\mathbf{x}_n})$, with

$$\mathbf{m}_{\mathbf{x}_n} = \beta_n \boldsymbol{\Sigma}_{\mathbf{x}_n} \mathbf{H}^\top \mathbf{y}_n, \quad \boldsymbol{\Sigma}_{\mathbf{x}_n}^{-1} = \beta_n \mathbf{H}^\top \mathbf{H} + \boldsymbol{\Theta}, \quad \boldsymbol{\Theta} = \mathbb{E}_{q(\boldsymbol{\omega})} [\text{diag}(\boldsymbol{\omega})].$$

Inference in the augmented model

$$p(\{\mathbf{x}_n\}, \mathbf{h}, \boldsymbol{\omega} \mid \{\mathbf{y}_n\}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \approx q(\{\mathbf{x}_n\}) q(\mathbf{h}) q(\boldsymbol{\omega})$$

- The filtered images are given by $q(\mathbf{x}_n) = \mathcal{N}(\mathbf{x}_n \mid \mathbf{m}_{\mathbf{x}_n}, \boldsymbol{\Sigma}_{\mathbf{x}_n})$, with

$$\mathbf{m}_{\mathbf{x}_n} = \beta_n \boldsymbol{\Sigma}_{\mathbf{x}_n} \mathbf{H}^\top \mathbf{y}_n, \quad \boldsymbol{\Sigma}_{\mathbf{x}_n}^{-1} = \beta_n \mathbf{H}^\top \mathbf{H} + \boldsymbol{\Theta}, \quad \boldsymbol{\Theta} = \mathbb{E}_{q(\boldsymbol{\omega})} [\text{diag}(\boldsymbol{\omega})].$$

- We estimate the blur as the mode of $q(\mathbf{h})$

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

\mathbf{C} and \mathbf{b} depend on $\mathbf{m}_{\mathbf{x}_n}$, $\boldsymbol{\Sigma}_{\mathbf{x}_n}$, and \mathbf{y}_n .

Inference in the augmented model

$$p(\{\mathbf{x}_n\}, \mathbf{h}, \boldsymbol{\omega} \mid \{\mathbf{y}_n\}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \approx q(\{\mathbf{x}_n\}) q(\mathbf{h}) q(\boldsymbol{\omega})$$

- The filtered images are given by $q(\mathbf{x}_n) = \mathcal{N}(\mathbf{x}_n \mid \mathbf{m}_{\mathbf{x}_n}, \boldsymbol{\Sigma}_{\mathbf{x}_n})$, with

$$\mathbf{m}_{\mathbf{x}_n} = \beta_n \boldsymbol{\Sigma}_{\mathbf{x}_n} \mathbf{H}^\top \mathbf{y}_n, \quad \boldsymbol{\Sigma}_{\mathbf{x}_n}^{-1} = \beta_n \mathbf{H}^\top \mathbf{H} + \boldsymbol{\Theta}, \quad \boldsymbol{\Theta} = \mathbb{E}_{q(\boldsymbol{\omega})} [\text{diag}(\boldsymbol{\omega})].$$

- We estimate the blur as the mode of $q(\mathbf{h})$

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

\mathbf{C} and \mathbf{b} depend on $\mathbf{m}_{\mathbf{x}_n}$, $\boldsymbol{\Sigma}_{\mathbf{x}_n}$, and \mathbf{y}_n .

- For $q(\boldsymbol{\omega})$ we don't need the full distribution, only its first moment. Its expression is a consequence of the GSM representation,

$$\mathbb{E}_{q(\omega_n^i)} [\omega_n^i] = \frac{\alpha_n \tanh(\alpha_n \xi_n^i)}{\xi_n^i}, \quad \xi_n^i = \sqrt{\mathbb{E}_{q(x_n^i)} [(x_n^i)^2]}.$$

Algorithm

1. Iterate through the previous updates to approximate the optimal variational distributions,

$$\begin{array}{ccccc} q^0(\{\mathbf{x}_n\}) & & q^1(\{\mathbf{x}_n\}) & & q^T(\{\mathbf{x}_n\}) \\ q^0(\mathbf{h}) & \longrightarrow & q^1(\mathbf{h}) & \longrightarrow \cdots \longrightarrow & q^T(\mathbf{h}) \\ q^0(\boldsymbol{\omega}) & & q^1(\boldsymbol{\omega}) & & q^T(\boldsymbol{\omega}) \end{array}$$

Algorithm

1. Iterate through the previous updates to approximate the optimal variational distributions,

$$\begin{array}{ccccc} q^0(\{\mathbf{x}_n\}) & & q^1(\{\mathbf{x}_n\}) & & q^T(\{\mathbf{x}_n\}) \\ q^0(\mathbf{h}) & \longrightarrow & q^1(\mathbf{h}) & \longrightarrow & \cdots \longrightarrow q^T(\mathbf{h}) \\ q^0(\boldsymbol{\omega}) & & q^1(\boldsymbol{\omega}) & & q^T(\boldsymbol{\omega}) \end{array}$$

2. The mode $\hat{\mathbf{h}} = \operatorname{argmax}_{\mathbf{h}} q^T(\mathbf{h})$ is used to estimate the latent clean image as

$$\hat{\mathbf{x}} = \operatorname{arg\,min}_{\mathbf{x}} \left\{ \frac{1}{2} \left\| \hat{\mathbf{h}} * \mathbf{x} - \mathbf{y} \right\|^2 + \lambda \sum_{n=1}^N \left\| \mathbf{F}_n \mathbf{x} \right\|^p \right\},$$

with $p = 0.8$.

Plan

1. Blind Image Deconvolution
2. The Hyperbolic Secant (HS) distribution
3. Modelling and inference
4. **Results**
5. Conclusions

Experimental framework

- We used the Levin data set.

Experimental framework

- We used the Levin data set.
- We compared the proposed method with
 1. Other analytical approaches based on SGs priors: log, ℓ_1 , MoG, exp, Huber SG, and ECP.
 2. Deep Learning-based methods: SelfDeblur and Li.

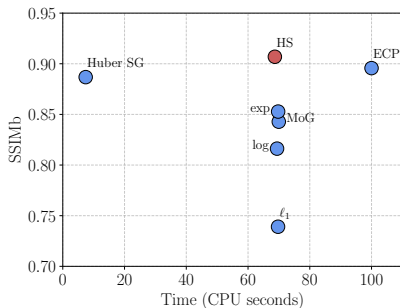
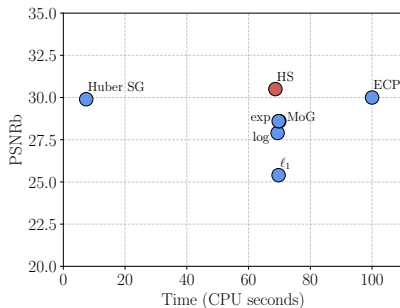
Experimental framework

- We used the Levin data set.
- We compared the proposed method with
 1. Other analytical approaches based on SGs priors: log, ℓ_1 , MoG, exp, Huber SG, and ECP.
 2. Deep Learning-based methods: SelfDeblur and Li.
- High-pass filters: first-order horizontal and vertical differences.

Experimental framework

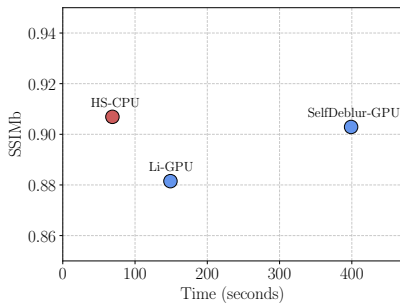
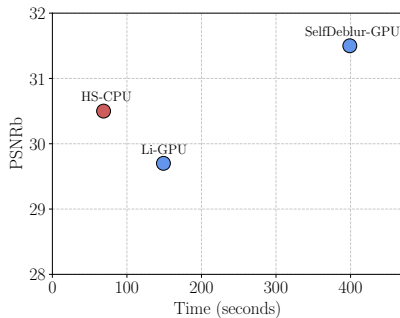
- We used the Levin data set.
- We compared the proposed method with
 1. Other analytical approaches based on SGs priors: log, ℓ_1 , MoG, exp, Huber SG, and ECP.
 2. Deep Learning-based methods: SelfDeblur and Li.
- High-pass filters: first-order horizontal and vertical differences.
- Hyperparameters obtained using grid search: $\beta = 10^4$, $\alpha_1 = 10^{2.4}$, and $\alpha_2 = 10^{2.15}$.

HS vs analytical approaches



- The proposed HS achieves the best results in terms of absolute performance.
- It is faster than the other methods except Huber SG.

HS vs Deep Learning approaches



- The proposed HS achieves competitive results in terms of absolute performance.
- It is faster than the other methods
- It does not need a GPU.

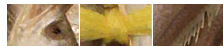
Comparison using real images



Observed



SelfDeblur



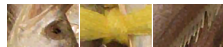
Li



HS (ours)



Huber SG



ECP

Plan

1. Blind Image Deconvolution
2. The Hyperbolic Secant (HS) distribution
3. Modelling and inference
4. Results
- 5. Conclusions**

Conclusions

- First use of the HS distribution in BID.

Conclusions

- First use of the HS distribution in BID.
- The GSM representation provides a new BID Bayesian method.

Conclusions

- First use of the HS distribution in BID.
- The GSM representation provides a new BID Bayesian method.
- Competitive or superior performance in the tested datasets.

Conclusions

- First use of the HS distribution in BID.
- The GSM representation provides a new BID Bayesian method.
- Competitive or superior performance in the tested datasets.
- Future work:
 - Automatic hyperparameter estimation.
 - Extensive evaluation on larger datasets.
 - Coupling the HS distribution with Deep Learning methods.

Thank you! ♡