Blind Deblurring with the MAP & Learnt Priors

Pierre Weiss, CNRS, Université de Toulouse

Joint work with Minh Hai Nguyen & Edouard Pauwels

Introduction

Notation & Problem

Assume that

$$y = A(\bar{\theta})\bar{x} + b$$

where:

- $A(\theta) : \mathbb{R}^N \to \mathbb{R}^M$ is a linear operator
- $b \in \mathbb{R}^M$: noise
- $\theta \in \mathbb{R}^P$: parameterization of the operator

Notation & Problem

Assume that

$$y = A(\bar{\theta})\bar{x} + b$$

where:

- $A(\theta) : \mathbb{R}^N \to \mathbb{R}^M$ is a linear operator
- $b \in \mathbb{R}^M$: noise
- $\theta \in \mathbb{R}^P$: parameterization of the operator

Our main example

$$A(\theta)x = h(\theta) \star x$$

$h(\theta)$: kernel depending on parameter θ (e.g. pixel values, Zernike coefficients)

h is called PSF (Point Spread Function)

Notation & Problem

Assume that

$$y = A(\bar{\theta})\bar{x} + b$$

where:

- $A(\theta) : \mathbb{R}^N \to \mathbb{R}^M$ is a linear operator
- $b \in \mathbb{R}^M$: noise
- $\theta \in \mathbb{R}^P$: parameterization of the operator

Our main example

$$A(\theta)x = h(\theta) \star x$$

$h(\theta)$: kernel depending on parameter θ (e.g. pixel values, Zernike coefficients)

h is called PSF (Point Spread Function)

Blind Inverse Problem

Find $(\hat{\theta}, \hat{x}) \approx (\bar{\theta}, \bar{x})$ from y.

- Looks much harder that classical linear inverse problems
- Yet... Dominant in applications

Posterior Maximization

Assuming that everything is random:

- x realization of $\boldsymbol{x} \sim p_x$ (learnt prior)
- θ realization of $\boldsymbol{\theta} \sim p_\theta$ uniform over a compact set Θ
- b realization of b $\sim \mathcal{N}(0, \sigma^2 I)$

MAP solution

.

$$
\begin{aligned}
(\hat{x}, \hat{\theta}) &= \underset{x \in \mathbb{R}^N, \theta \in \Theta}{\operatorname{argmax}} \; p(x, \theta | y) \\
&= \underset{x, \theta \in \Theta}{\operatorname{argmin}} \; \frac{1}{2\sigma^2} \|A(\theta)x - y\|_2^2 - \log p_x(x) \\
&= \underset{x, \theta \in \Theta}{\operatorname{argmin}} \; \frac{1}{2\sigma^2} \|A(\theta)x - y\|_2^2 + g(x)
\end{aligned}
$$

with $g(x) = -\log p_x(x)$

A Very Popular Framework

About 5-10 papers/year since last 20 years including:

- T. Chan et al. Total Variation Blind Deconvolution, IEEE IP 1998
- A. Levin et al. Understanding blind deconvolution, CVPR 2009
- D. Perrone et al. Total variation blind deconvolution, CVPR 2014
- H. Chung et al. Parallel diffusion models, CVPR 2023
- C. Laroche et al. Fast diffusion EM, WACV 2024

But... It Doesn't Work!

Attempt to solve the problem with state-of-the-art diffusion prior

Main Results

A Preliminary Observation...

Noiseless setting:

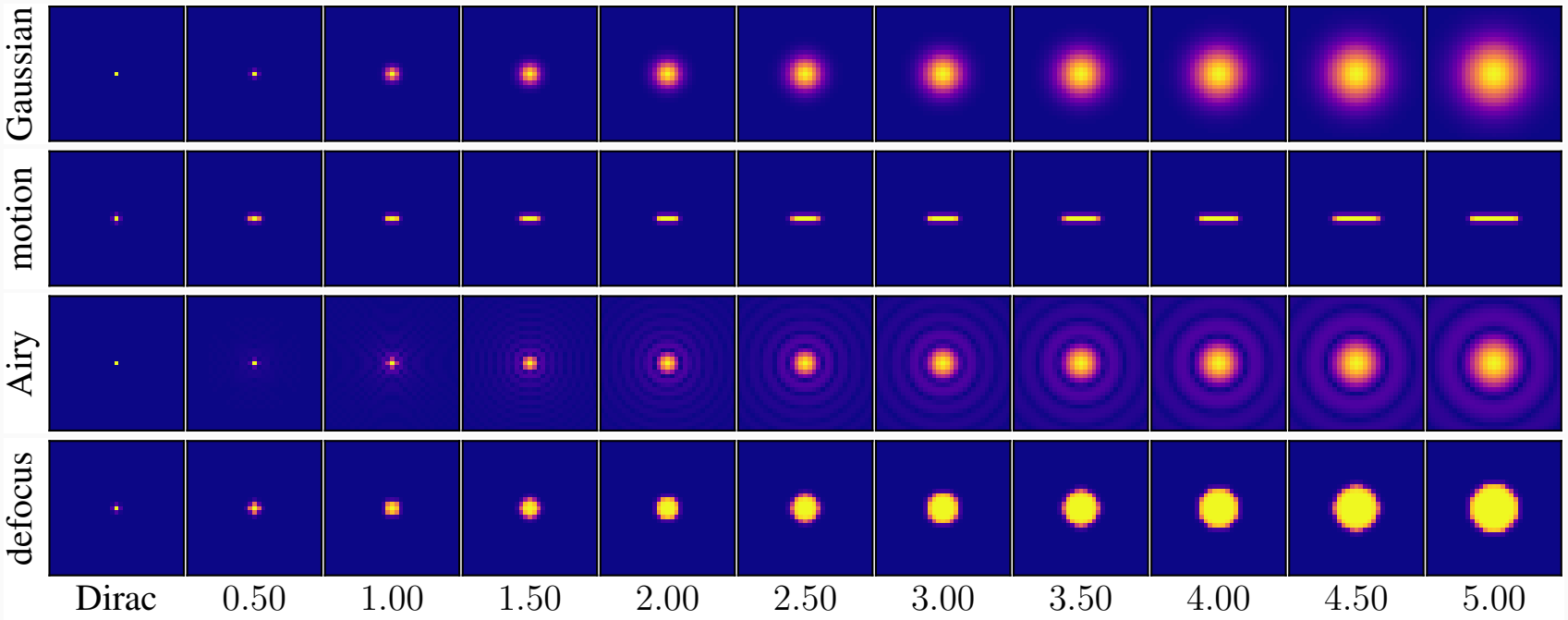
$$y = h(\theta) \star x$$

Consider the simplex constrained alternate minimization:

$$\begin{aligned} x_0 &= y \\ h_{k+1} &= \underset{h \in \Delta_{N-\mathbf{1}}}{\operatorname{argmin}} \frac{1}{2\sigma^2} \|h \star x_k - y\|_2^2 \\ x_{k+1} &= \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2\sigma^2} \|h_{k+1} \star x - y\|_2^2 - \log p_x(x) \end{aligned}$$

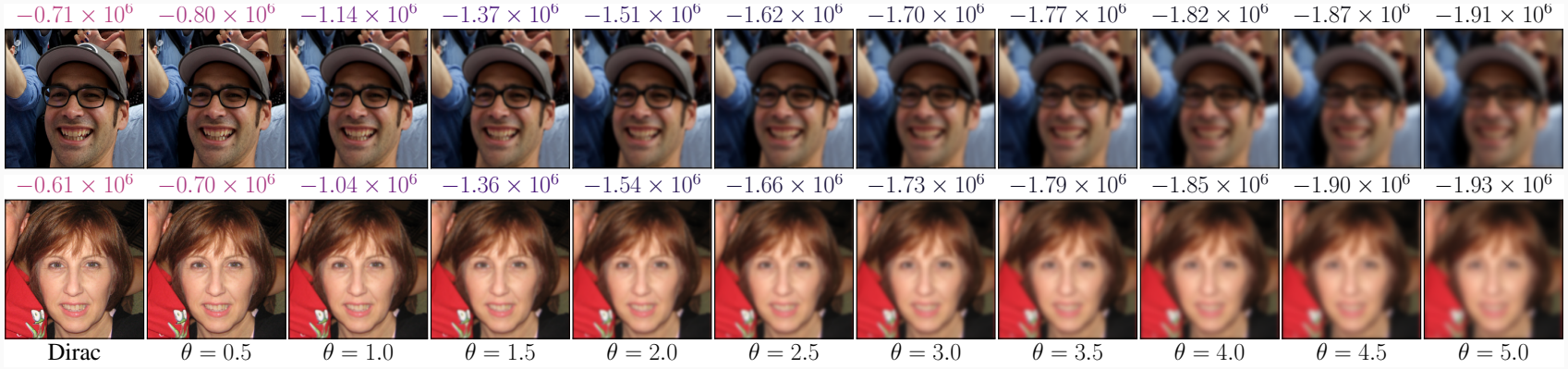
Then $h_1 = \delta$ and if y is likely, we are stuck to (δ, y) !

Blurry Images Are Likely?



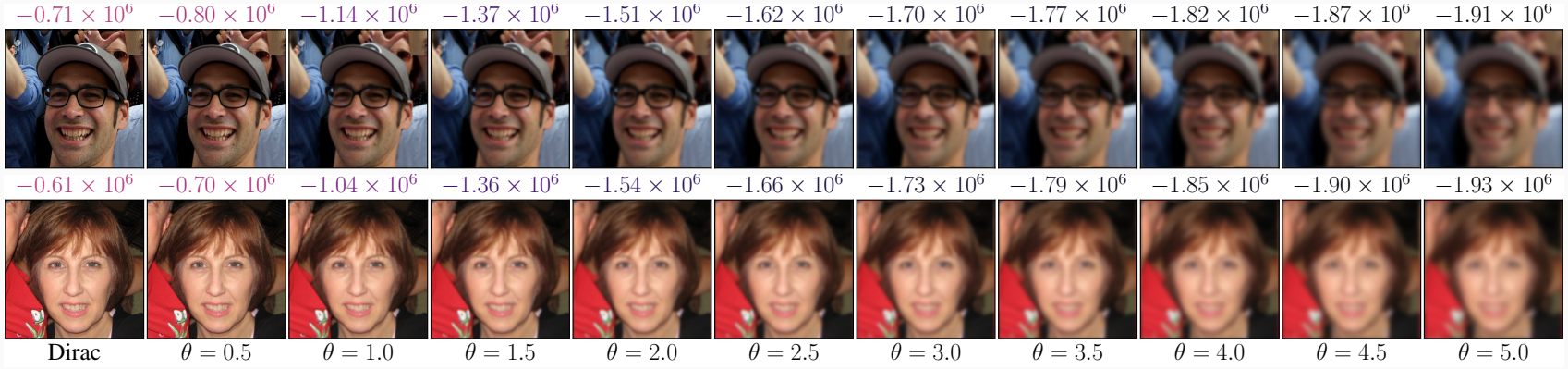
A 1D family of PSFs

Blurry Images Are Likely?

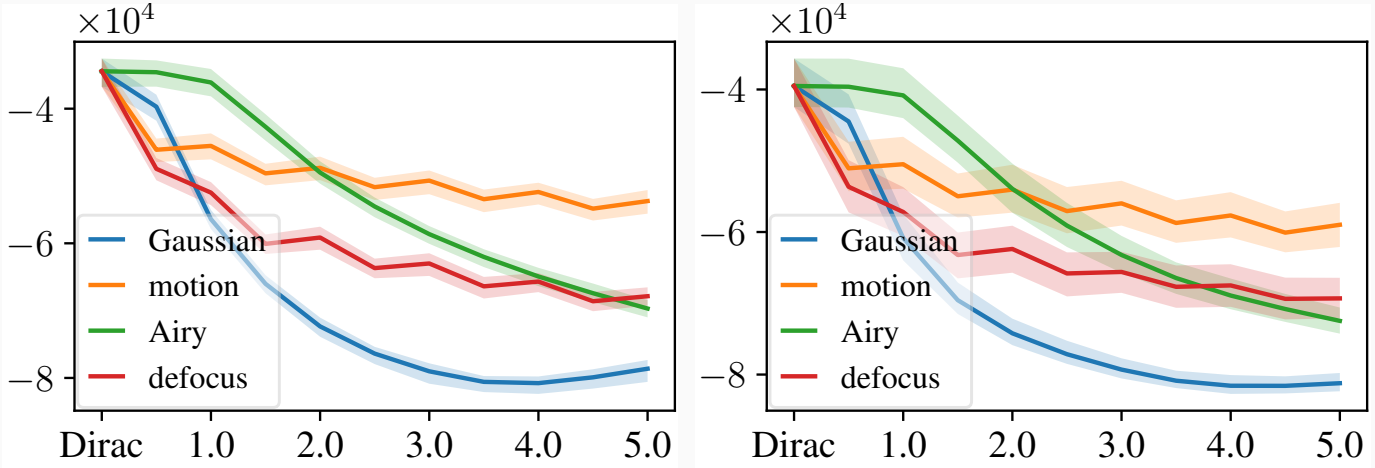


Potential g w.r.t. PSF width

Blurry Images Are Likely?



Potential g w.r.t. PSF width



Replicated over 100 images with two learnt diffusion priors (Karras EDM)
Note: evaluating g = ODE integration = heavy

Explaining the MAP Failure

Theorem (The global minimizer is bad)

Let $\mathcal{H} = \{h(\theta), \theta \in \Theta\}$ *and* $\delta \in \mathcal{H}$.

If more blurry = more likely, that is $\forall x, h \in \mathcal{H}$*:*

$$g(h \star x) \leq g(x) \tag{1}$$

then

$$(\hat{x}_{MAP}^{denoise}, \delta) \in \underset{x \in \mathbb{R}^N, h \in \mathcal{H}}{\operatorname{argmin}} \frac{1}{2\sigma^2} \|h \star x - y\|_2^2 + g(x)$$

with

$$\hat{x}_{MAP}^{denoise} = \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2\sigma^2} \|x - y\|_2^2 + g(x)$$

[1] A. Levin et al. Understanding blind deconvolution, CVPR 2009
[2] D. Perrone et al. Total variation blind deconvolution, CVPR 2014

Explaining the MAP Failure

Theorem (The global minimizer is bad)

Let $\mathcal{H} = \{h(\theta), \theta \in \Theta\}$ *and* $\delta \in \mathcal{H}$.

If more blurry = more likely, that is $\forall x, h \in \mathcal{H}$:

$$g(h \star x) \leq g(x) \tag{1}$$

then

$$(\hat{x}_{MAP}^{denoise}, \delta) \in \underset{x \in \mathbb{R}^N, h \in \mathcal{H}}{\operatorname{argmin}} \frac{1}{2\sigma^2} \|h \star x - y\|_2^2 + g(x)$$

with

$$\hat{x}_{MAP}^{denoise} = \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2\sigma^2} \|x - y\|_2^2 + g(x)$$

Remarks

- This fact is well known for handcrafted priors (Hölder) [1] [2]
- We dreamt of a different behavior for learnt priors
- Learnt priors = Tweedie = MMSE = averaging / blur

[1] A. Levin et al. Understanding blind deconvolution, CVPR 2009
[2] D. Perrone et al. Total variation blind deconvolution, CVPR 2014

Why so many people do it? Even in the recent past?

Sufficient Recovery Conditions for Non Blind Inverse Problems

Nonblind, noiseless setting:

$$y = A\bar{x} \qquad \text{and} \qquad F(x) = g(x) + \frac{1}{2\sigma^2}\|Ax - y\|_2^2$$

Assume that $g(x) = -\log p_x(x)$ is C^2

Theorem (A compressed sensing type observation)

The propositions below are equivalent:

1. $\bar{x}$ = *strict local minimizer of* F
2. $\bar{x}$ *is a second order critical point (SOCP) of* g *(not a saddle)*

$$\nabla g(\bar{x}) = 0$$
$$\nabla^2 g(\bar{x}) \succeq 0$$

and

$$\ker(\nabla^2 g(\bar{x})) \cap \ker(A) = \{0\}$$

Points recovered by MAP&learnt priors = SOCP of $-\log p_x$!

Recovery Conditions for Blind Inverse Problems

Assume

$$\nabla g(\bar{x}) = 0, \quad \nabla^2 g(\bar{x}) \succeq 0, \quad A(\bar{\theta})\bar{x} = \bar{y}$$

Set

$$J(\theta) = \frac{\partial}{\partial \theta} A(\theta)\bar{x} \in \mathbb{R}^{M \times P}$$

Recovery Conditions for Blind Inverse Problems

Assume

$$\nabla g(\bar{x}) = 0, \quad \nabla^2 g(\bar{x}) \succeq 0, \quad A(\bar{\theta})\bar{x} = \bar{y}$$

Set

$$J(\theta) = \frac{\partial}{\partial \theta} A(\theta)\bar{x} \in \mathbb{R}^{M \times P}$$

Theorem (Extension to blind inverse problems)

Set

$$F(x, \theta) = \frac{1}{2\sigma^2} \|A(\theta)\bar{x} - \bar{y}\|_2^2 + g(x)$$

If

$$\ker(\nabla^2 g(\bar{x})) \cap \ker(A(\bar{\theta})) = \{0\}$$
$$\ker(J(\bar{\theta})) = \{0\}$$
$$A(\bar{\theta}) \ker(\nabla^2 g(\bar{x})) \cap \operatorname{ran}(J(\bar{\theta})) = \{0\}$$

Then $(\bar{x}, \bar{\theta})$ *is a strict local minimizer of* F

Recovery Conditions for Blind Inverse Problems

Assume

$$\nabla g(\bar{x}) = 0, \quad \nabla^2 g(\bar{x}) \succeq 0, \quad A(\bar{\theta})\bar{x} = \bar{y}$$

Set

$$J(\theta) = \frac{\partial}{\partial \theta} A(\theta)\bar{x} \in \mathbb{R}^{M \times P}$$

Theorem (Extension to blind inverse problems)

Set

$$F(x, \theta) = \frac{1}{2\sigma^2} \|A(\theta)\bar{x} - \bar{y}\|_2^2 + g(x)$$

If

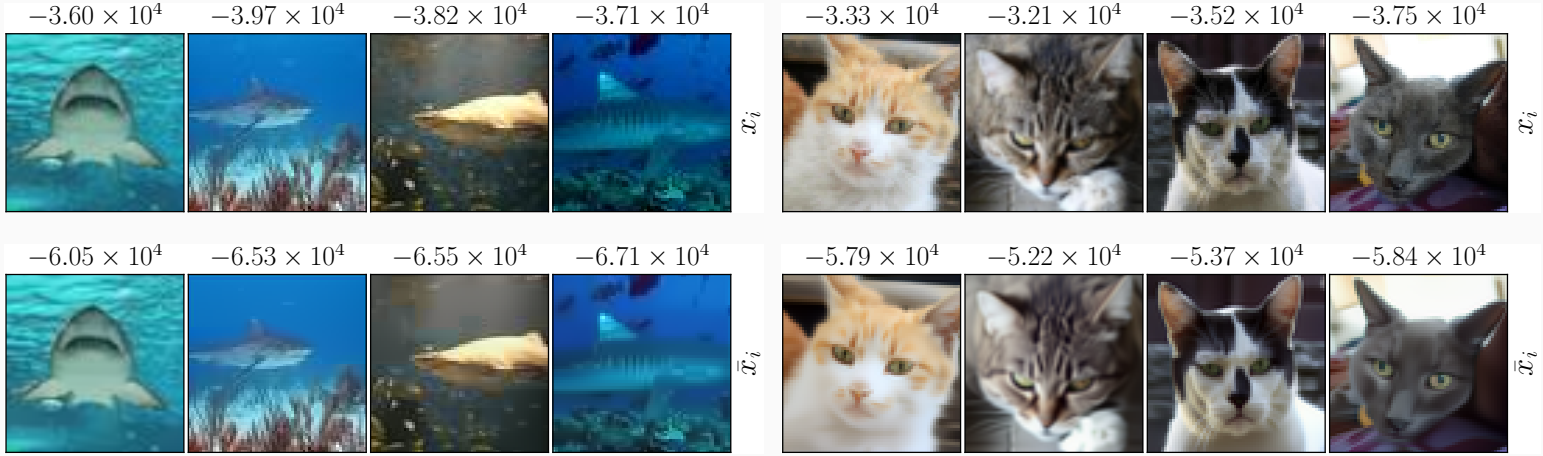
$$\ker(\nabla^2 g(\bar{x})) \cap \ker(A(\bar{\theta})) = \{0\}$$
$$\ker(J(\bar{\theta})) = \{0\}$$
$$A(\bar{\theta}) \ker(\nabla^2 g(\bar{x})) \cap \operatorname{ran}(J(\bar{\theta})) = \{0\}$$

Then $(\bar{x}, \bar{\theta})$ *is a strict local minimizer of* F

Moreover, the result is stable to noise on y

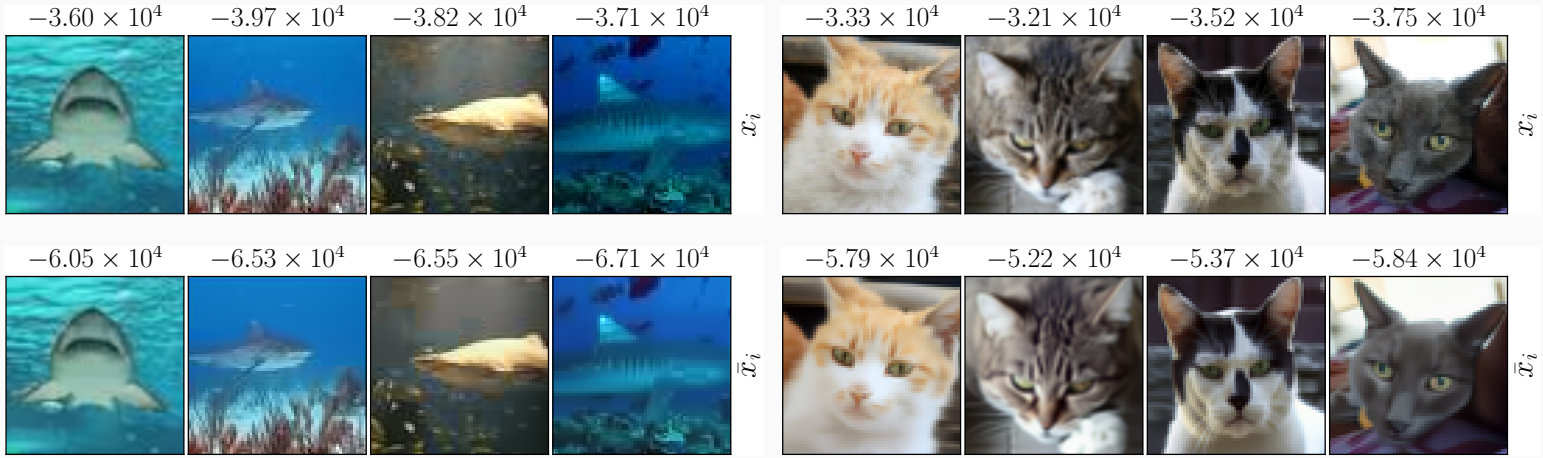
Second Order Critical Points of Learnt priors

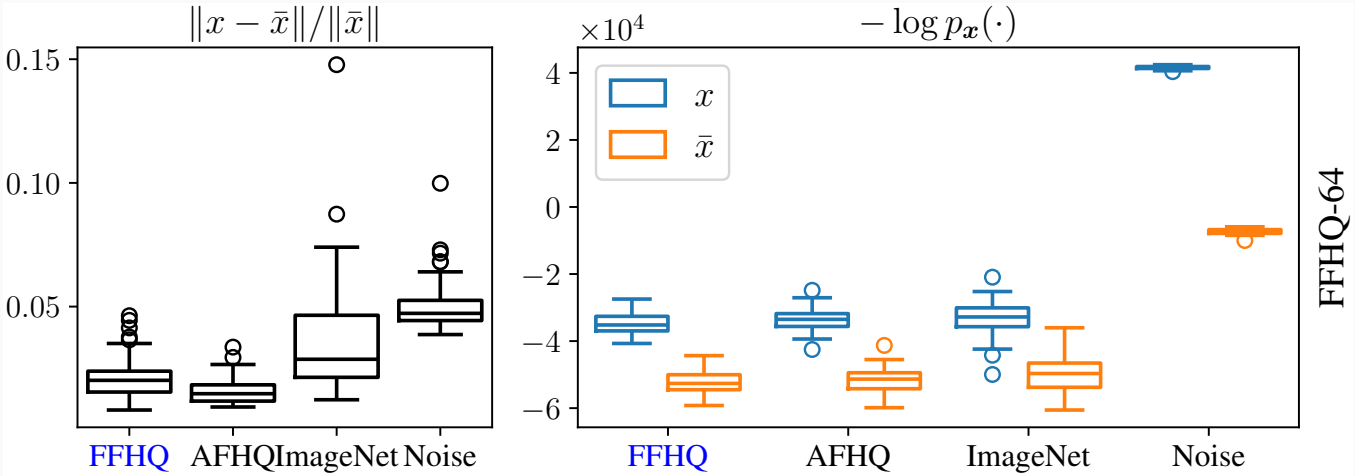
SOCP: take a collection (x_i), run a gradient descent on g, land at $(\bar{x}_i)$



Second Order Critical Points of Learnt priors

SOCP: take a collection (x_i), run a gradient descent on g, land at $(\bar{x}_i)$





SOCP lie near every point in space for EDM!

Erratic landscape, with a "deep valley" around an image manifold

Temporary Conclusions for Learnt Priors

$$F(x, \theta) = \frac{1}{2\sigma^2} \|A(\theta)\bar{x} - \bar{y}\|_2^2 + g(x)$$

- MAP = global minimizer = blurry image
  - This is because blurry image are more likely

Temporary Conclusions for Learnt Priors

$$F(x,\theta) = \frac{1}{2\sigma^2}\|A(\theta)\bar{x} - \bar{y}\|_2^2 + g(x)$$

- MAP = global minimizer = blurry image
  - This is because blurry image are more likely
- Posterior possesses local minimizers around SOCP

Temporary Conclusions for Learnt Priors

$$F(x,\theta) = \frac{1}{2\sigma^2}\|A(\theta)\bar{x} - \bar{y}\|_2^2 + g(x)$$

- MAP = global minimizer = blurry image
  - This is because blurry image are more likely
- Posterior possesses local minimizers around SOCP
- SOCP are:
  - Nearly everywhere in space
  - Valuable SOCPs around natural images

Temporary Conclusions for Learnt Priors

$$F(x, \theta) = \frac{1}{2\sigma^2} \|A(\theta)\bar{x} - \bar{y}\|_2^2 + g(x)$$

- MAP = global minimizer = blurry image
  - This is because blurry image are more likely
- Posterior possesses local minimizers around SOCP
- SOCP are:
  - Nearly everywhere in space
  - Valuable SOCPs around natural images

Can we design algorithms converging locally?

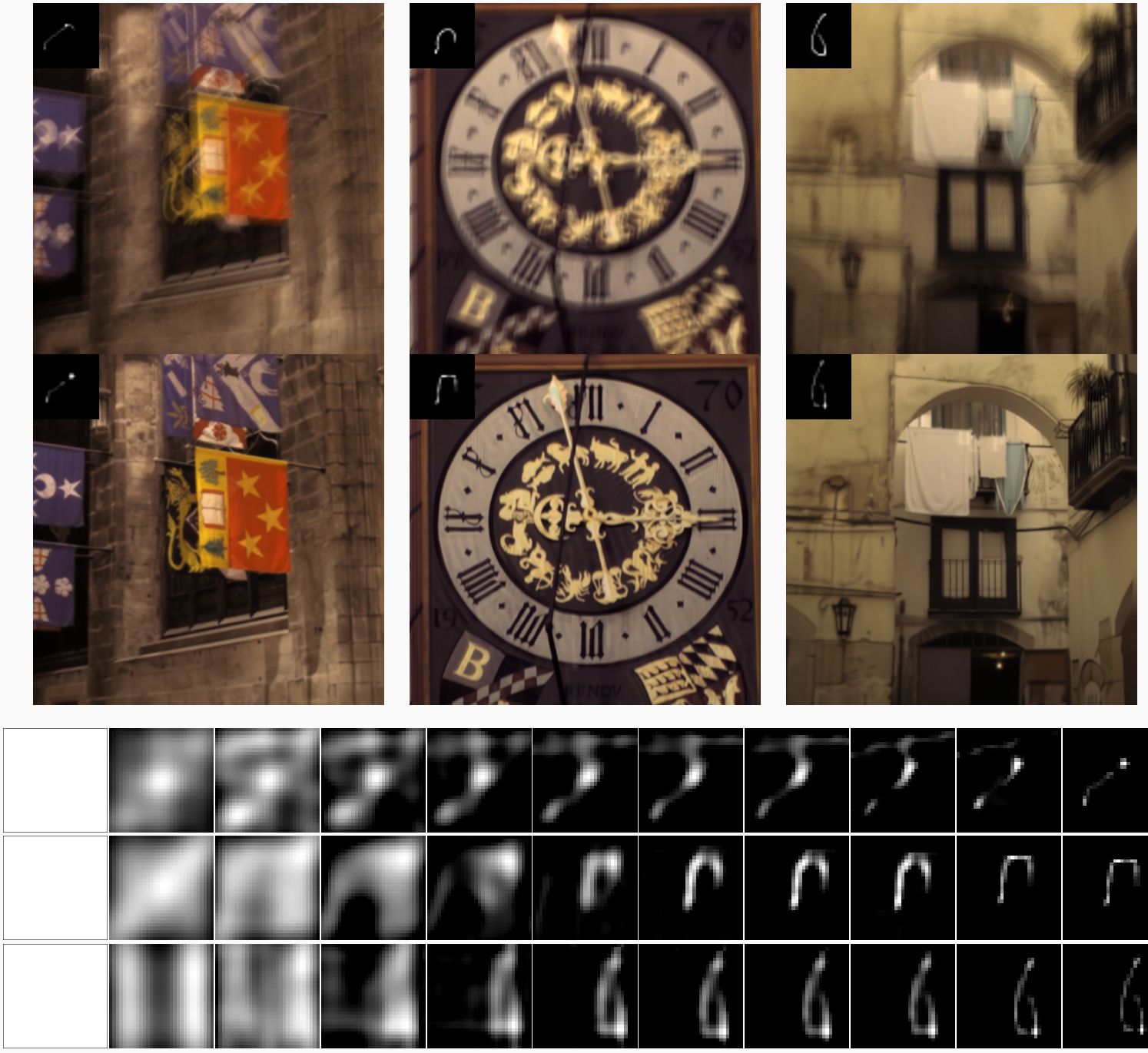
A simple algorithm

We consider the following algorithm:

- Set a parameter θ_0 corresponding to a large PSF
- $x_{k+1} = \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} F(x, \theta_k)$ (e.g. proximal gradient descent)
- $\theta_{k+1} = \theta_k - \tau_k \nabla_\theta F(x_{k+1}, \theta_k)$

That is:

- Exact minimization in x
- One-step gradient descent in θ



A pleasant evolution (parameterization in space domain)

Diffraction limited kernels

It now works fine...

Take Home

M.H. Nguyen

E. Pauwels

- Posterior possesses good local minimizers
- Heuristic methods can end up there (sometimes)

Take Home

M.H. Nguyen

E. Pauwels

- Posterior possesses good local minimizers
- Heuristic methods can end up there (sometimes)
- Still... Slow and unreliable
- Other methods like MMSE should likely be preferred