Equivariant Denoisers for Image Restoration

Workshop - Advances in learning-based image restoration

M. Renaud[1, 2] A. Leclaire[2] N. Papadakis[1]

[1]IMB
Institut de Mathématiques
de Bordeaux

[2]LTCI
Télécom Paris

[3]Université Paris-Cité
MAP5

December 09 2024

Plan

Introduction to inverse problem in imaging

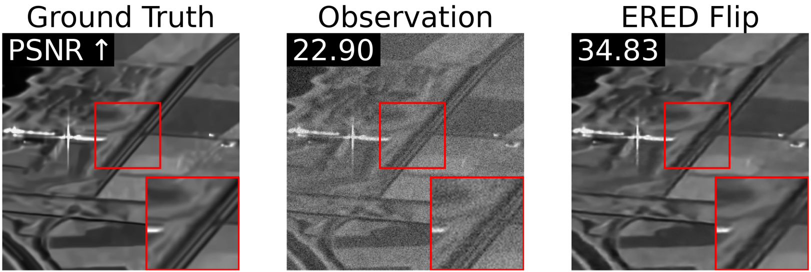

Figure – Image despeckling with a number of looks of 50 with ERED algorithm (random flip)

Introduction to inverse problem

Physical model of the degradation

$$y \sim \mathcal{N}(\mathcal{A}(x)),$$

with

- $x \in \mathbb{R}^d$: clean image
- $y \in \mathbb{R}^m$: noisy observation
- $\mathcal{A} : \mathbb{R}^d \to \mathbb{R}^m$: degradation operator
- $\mathcal{N}$: law of noise model

Linear system :

$$y = Ax + n,$$

with $A \in \mathbb{R}^{m \times d}$ and $n \sim \mathcal{N}(0, \sigma_y^2)$.

How to recover x from the observation y ?

Introduction to maximum a posteriori

We look for

$$\begin{aligned}\arg\max_{x\in\mathbb{R}^d} p(x|y) &= \arg\max_{x\in\mathbb{R}^d} \frac{p(y|x)p(x)}{p(y)} \\ &= \arg\min_{x\in\mathbb{R}^d} \underbrace{-\log p(y|x)}_{f(x)} \underbrace{-\log p(x)}_{r(x)}\end{aligned}$$

We want to solve the problem

$$\arg\min_{x\in\mathbb{R}^d} \mathcal{F}(x) := f(x) + \lambda r(x).$$

Gradient descent scheme :

$$x_{k+1} = x_k - \delta\left(\nabla f(x_k) + \lambda \nabla r(x_k)\right)$$

Regularisation by denoising (RED)

To solve the previous problem, we train a Deep Neural Network D_σ to denoise images. Thanks to the Tweedie formula, we make the following approximations

$$\nabla r(x) = -\nabla \log p(x) \approx -\nabla \log p_\sigma(x) = \frac{1}{\sigma^2}\left(x - D^\star_\sigma(x)\right) \approx \frac{1}{\sigma^2}\left(x - D_\sigma(x)\right).$$

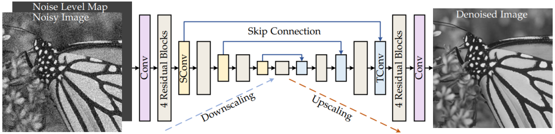

Figure – Drunet denoiser [1]

1. Kai Zhang et al., Plug-and-Play Image Restoration with Deep Denoiser Prior, 2021

Regularisation by denoising (RED) [2]

Algorithm RED

Param. : init. $x_0 \in \mathbb{R}^d$, $\sigma > 0$, $\lambda > 0$, $\delta > 0$, $N \in \mathbb{N}$
Input : degraded image y
Output : restored image x_N
for $k = 0, 1, \ldots, N-1$ **do**
    $x_{k+1} \leftarrow x_k - \delta \nabla f(x_k, y) - \frac{\lambda\delta}{\sigma^2}(x_k - D_\sigma(x_k))$
end for

2. Y. Romano, M. Elad, and P. Milanfar. The little engine that could : Regularization by denoising (RED). SIAM Journal on Imaging Sciences, 2017.

Stochastic deNOising REgularization (SNORE) [3]

Algorithm SNORE

Param. : init. $x_0 \in \mathbb{R}^d$, $\sigma > 0$, $\lambda > 0$, $\delta > 0$, $N \in \mathbb{N}$
Input : degraded image y
Output : restored image x_N
for $k = 0, 1, \dots, N-1$ **do**
    $\tilde{x}_k \leftarrow x_k + \sigma \epsilon_k$ with $\epsilon_k \leftarrow \mathcal{N}(0, I_d)$
    $x_{k+1} \leftarrow x_k - \delta \nabla f(x_k, y) - \frac{\lambda\delta}{\sigma^2}(x_k - D_\sigma(\tilde{x}_k))$
end for

3. M. Renaud, J. Prost, A. Leclaire, and N. Papadakis. Plug-and-play image restoration with stochastic denoising regularization, ICML, 2024.

Equivariant Regularization by Denoising (eq. RED) [4]

Algorithm eq. RED

Param. : init. $x_0 \in \mathbb{R}^d$, $\sigma > 0$, $\lambda > 0$, $\delta > 0$, $N \in \mathbb{N}$
Input : degraded image y
Output : restored image x_N
for $k = 0, 1, \ldots, N-1$ **do**
  $g_k \sim \mathcal{G}$
  $x_{k+1} \leftarrow x_k - \delta \nabla f(x_k, y) - \frac{\lambda\delta}{\sigma^2}\left(x_k - g_k^{-1} D_\sigma(g_k x_k)\right)$
end for

g_k are linear invertible transformations.

4. M. Terris, T. Moreau, N. Pustelnik, and J. Tachella. Equivariant plug-and-play image reconstruction. ArXiv, 2024.

Plan

Some notations

- $\mathcal{G}$ is a set of differentiable transformations of $\mathbb{R}^d$.
- $g : \mathbb{R}^d \to \mathbb{R}^d$ denote a transformation of $\mathcal{G}$.
- G is a random variable of law π on $\mathcal{G}$.

Example of invariance

Figure – Set of rotated images of an image.

Unified formulation of ERED

Algorithm ERED

1: **Param. :** init. $x_0 \in \mathbb{R}^d$, $\sigma > 0$, $\lambda > 0$, $\delta > 0$, $N \in \mathbb{N}$
2: **Input :** degraded image y
3: **Output :** restored image x_N
4: **for** $k = 0, 1, \ldots, N-1$ **do**
5: $\quad G \sim \pi$
6: $\quad x_{k+1} \leftarrow x_k - \delta \nabla f(x_k, y) - \frac{\lambda\delta}{\sigma^2}\left(J_G^T(x_k)G(x_k) - J_G^T(x_k)D_\sigma(G(x_k))\right)$
7: **end for**

- RED : $\mathcal{G} = \{I_d\}$.
- eq. RED : $\mathcal{G}$ is a finite set of linear isometries and π uniform, then $J_G^T(x) = G^{-1}$.
- SNORE : $\mathcal{G}$ is a set of translation, $g_z(x) = x + \sigma z$, and $\pi(g_z) = \mathcal{N}(z; 0, \sigma^2 I_d)$, then $J_G^T(x) = I_d$.

Notion of invariance and π-equivariance

Definition (Invariance)

A density p on $\mathbb{R}^d$ is said to be invariant to a set of transformations $\mathcal{G}$ if $\forall g \in \mathcal{G}$, $p = p \circ g$ a.e.

Definition (π-equivariance)

A density p on $\mathbb{R}^d$ is said to be π-equivariant if $\mathbb{E}_{G\sim\pi}[|\log(p \circ G)|] < \infty$ and $\log p = \mathbb{E}_{G\sim\pi}\left[\log(p \circ G)\right]$.

Why does the Jacobian appear ?

For $p \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}_+^*)$ and $g \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^d)$, we have

$$\nabla \log(p \circ g)(x) = \frac{\nabla(p \circ g)(x)}{(p \circ g)(x)} = \frac{J_g^T(x)\nabla p(g(x))}{(p \circ g)(x)} = J_g^T(x)(\nabla \log p)(g(x)),$$

with $x \in \mathbb{R}^d$.

The π-equivariant regularization

We introduce the π-equivariant regularization by

$$r_\sigma^\pi(x) := -\mathbb{E}_{G\sim\pi}\left(\log(p_\sigma \circ G)(x)\right)$$
$$s_\sigma^\pi(x) := -\mathbb{E}_{G\sim\pi}\left(J_G^T(x)(\nabla \log p_\sigma)(G(x))\right).$$

Thanks to the Tweedie formula, we get

$$s_\sigma^\pi(x) = \frac{1}{\sigma^2}\left(\mathbb{E}_\pi\left[J_G^T(x)G(x)\right] - \tilde{D}_\sigma^*(x)\right) \approx \frac{1}{\sigma^2}\left(\mathbb{E}_\pi\left[J_G^T(x)G(x)\right] - \tilde{D}_\sigma(x)\right),$$

with $\tilde{D}_\sigma =$ the equivariant denoiser.
ERED is a stochastic gradient descent to solve

$$\underset{x\in\mathbb{R}^d}{\arg\min}\,\mathcal{F}(x) = f(x) + \lambda r_\sigma^\pi(x)$$

The π-equivariant regularization is π-equivariant

Proposition

If $\mathcal{G}$ is a compact Hausdorff topological group and π the associated right-invariant Haar measure, then r_σ^π is π-equivariant.

Plan

With the exact MMSE Denoiser D_σ^*

We define the set of critical points $S_\sigma = \{x \in \mathbb{R}^d | \nabla \mathcal{F}(x) = 0\}$

$$\Lambda_K = \bigcap_{k \in \mathbb{N}} \{x_k \in K\}.$$

Proposition

Let $(x_k)_{k\in\mathbb{N}}$ be the iterates generated by ERED with the exact MMSE Denoiser D_σ^*. Then, under Assumptions, we have almost surely on Λ_K

$$\lim_{k\to+\infty} d(x_k, S_\sigma) = 0, \tag{1}$$

$$\lim_{k\to+\infty} \|\nabla \mathcal{F}(x_k)\| = 0, \tag{2}$$

and $(\mathcal{F}(x_k))_{k\in\mathbb{N}}$ converges to a value of $\mathcal{F}(S_\sigma)$.

Plan

Qualitative experimental results

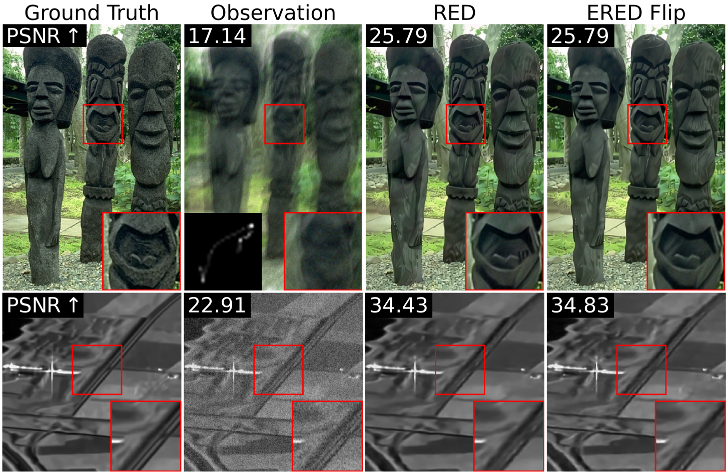

Figure – Deblurring (a motion blur kernel with input noise level $\sigma_y = 5/255$) and despeckling (number of looks 50)

Quantitave results for deblurring

Method	PSNR↑	SSIM↑	N↓
RED	32.25	0.84	400
ERED rotation	<u>32.53</u>	0.85	400
ERED translation	32.44	0.85	400
ERED flip	32.51	0.85	400
ERED subpixel rotation	32.32	0.85	400
ERED all transformations	31.94	0.83	400
SNORE	32.45	<u>0.86</u>	1000
Annealed SNORE	**32.89**	**0.87**	1500

Table – Quantitative comparison of image deblurring methods on 10 images from CBSD68 dataset with 10 different blur kernels (fixed and motion kernel of blur) and a noise level $\sigma_y = 5/255$.

Plan

Conclusion

Conclusion

- Unified formulation of ERED and geometrical interpretation
- Convergence results for ERED

Perspective

- Generalize to Equivariant PnP or Equivariant PnP-ULA
- Explore more geometrical transformations

Assumptions

Assumption

(a) *The step-size decreases to zero but not too fast :* $\sum_{k=0}^{+\infty} \delta_k = +\infty$ *and* $\sum_{k=0}^{+\infty} \delta_k^2 < +\infty$.
(b) *The data-fidelity term* $f : x \in \mathbb{R}^d \mapsto f(x) \in \mathbb{R}$ *is* $\mathcal{C}^\infty$.
(c) *The noisy prior score is sub-polynomial, i.e. there exist* $B \in \mathbb{R}^+$, $\beta \in \mathbb{R}$ *and* $n_1 \in \mathbb{N}$ *such that* $\forall \sigma > 0$, $\forall x \in \mathbb{R}^d$, $\|\nabla \log p_\sigma(x)\| \leq B\sigma^\beta(1 + \|x\|^{n_1})$.

Assumptions

Assumption

(a) *The random variable* J_G *has a uniform finite moment, i.e.* $\exists \epsilon > 0, M_{2+\epsilon} \geq 0$ *such that* $\forall x \in \mathbb{R}^d, \mathbb{E}_{G\sim\pi}(|||J_G(x)|||^{2+\epsilon}) \leq M_{2+\epsilon} < +\infty$.
(b) *The transformation has bounded moments on any compact, i.e.* $\forall \mathrm{K} \subset \mathbb{R}^d$ *compact,* $\forall m \in \mathbb{N}$, $\exists C_{\mathrm{K},m} < +\infty$ *such that* $\forall x \in \mathrm{K}, \mathbb{E}_{G\sim\pi}(\|G(x)\|^m) \leq C_{\mathrm{K},m}$.

Assumptions with D_σ

Assumption

The realistic denoiser D_σ is sub-polynomial, i.e. $\exists C > 0$ and $n_2 \in N$ such that $\forall x \in \mathbb{R}^d$, $\|D_\sigma(x)\| \leq C(1 + \|x\|^{n_2})$.

Assumption

For every compact K*, there exists C_K, such that* $\forall x \in \mathrm{K}, \forall g \in \mathcal{G}, \|g(x)\| \leq C_\mathrm{K}$.

Result with inexact denoiser D_σ

Proposition

Let $(x_k)_{k\in\mathbb{N}}$ be the sequence provided by ERED with an inexact denoiser D_σ. Then, under Assumptions, there exists M_K such that, almost surely on Λ_K :

$$\limsup_{k\to\infty} \|\nabla \mathcal{F}_\sigma^\pi(x_k)\| \leq M_\mathsf{K}\eta^{\frac{1}{2}} \tag{3}$$

$$\limsup_{k\to\infty} \mathcal{F}_\sigma^\pi(x_k) - \liminf_{k\to\infty} \mathcal{F}_\sigma^\pi(x_k) \leq M_\mathsf{K}\eta, \tag{4}$$

with the asymptotic bias $\eta = \limsup_{k\to\infty} \|\mathbb{E}(\xi_k)\|$.
Moreover, under last Assumption, we have

$$\eta \leq \frac{\lambda}{\sigma^2} \sup_{x\in\mathsf{K}} \mathbb{E}\left(|||J_G(x)|||\right) \|D_\sigma - D_\sigma^*\|_{\infty,\mathsf{L}}, \tag{5}$$

with $\mathsf{L} = \mathcal{B}(0, C_\mathsf{K})$.

Assumptions for critical points analysis

Assumption

(a) *The prior distribution* $p \in \mathrm{C}^1(\mathbb{R}^d,]0, +\infty[)$ *with* $\|p\|_\infty + \|\nabla p\|_\infty < +\infty$.
(b) J_G *has finite first moment, i.e.* $\sup_{x\in\mathbb{R}^d} \mathbb{E}_{G\sim\pi}(|||J_G(x)|||) < +\infty$.

Assumption

The data-fidelity term is continuously differentiable, i.e. $f \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R})$.

Critical point analysis

$$\mathsf{S} = \{x \in \mathbb{R}^d | \exists \sigma_n > 0 \text{ decreasing to } 0, x_n \in \mathsf{S}_{\sigma_n} \text{ such that } x_n \xrightarrow[n\to\infty]{} x\}.$$

Proposition

Under Assumptions, if the prior p is π-equivariant, we have

$$\mathsf{S} \subset \mathsf{S}^*.$$

Denoising performance

Denoising method	PNSR, $\sigma = 5/255$	PNSR $\sigma = 10/255$	PNSR $\sigma = 20/255$
Simple denoising	40.54	36.46	32.73
Rotation denoising	40.58	36.49	32.76
Translation denoising	40.53	36.44	32.71
Subpixel Rotation denoising	40.34	36.26	32.56
Flip denoising	40.58	36.49	32.76

Table – Denoising results on the CBSD68 dataset with various level of noise with the GS-DRUNet denoiser.

Deblurring with various denoiser

Denoiser	Restoration method	PNSR ↑	SSIM ↑
GS-DRUNet ($\sigma_y = \frac{5}{255}$)	RED	32.25	0.84
	ERED rotation	32.53	0.85
	ERED flip	32.51	0.85
DRUNet ($\sigma_y = \frac{5}{255}$)	RED	29.24	0.81
	ERED rotation	29.48	0.83
	ERED flip	29.44	0.82
DnCNN ($\sigma_y = \frac{1}{255}$)	RED	35.26	0.94
	ERED rotation	35.34	0.94
	ERED flip	35.32	0.94

Table – Deblurring results on CBSD10 (10 images extracted from CBSD68 dataset) with 10 kernels of blur (including fixed and motion blur) with different pre-trained denoisers.